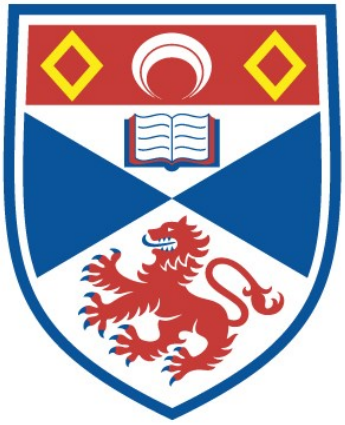
NOVEL SCANNING TECHNIQUES FOR CCD IMAGE CAPTURE AND DISPLAY

M.D. (Matthew David) Emberson

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1995

Novel Scanning Techniques for CCD Image Capture and Display

A thesis presented by
M. D. Emberson B.Sc. *Hons St. Andrews*
to the University of St. Andrews
in application for the degree of
Doctor of Philosophy.

August 24, 1994

To Eleanor ...

"Searching for that ultimate adrenaline rush, we prepare
to hurl ourselves over the brink! What fate awaits us?"

- Bill Watterson 1991.

My thanks to Dr. R.C.G. Killean, Dr. U. Arndt
and Prof. A. Mathieson.

Set in LaTeX using 12pt. Computer Modern.
Graphics in Adobe Illustrator® and Adobe Photoshop®.
Plots in Unigraph 2000.
Schematics in Viewlogic Workview.
Printed at 600dpi.

I *Matthew David Emberson* hereby certify that this thesis has been composed by myself, that it is a record of my own work, and that is has not been accepted in partial or complete fulfilment of any other degree or professional qualification.

Signed ..... Date 25th Aug. '94

I was admitted to the Faculty of Science of the University of St. Andrews under Ordinance General No 12 on *1st October 1988* and as a candidate for the degree of Ph.D. on *1st October 1989.*

Signed ..... Date 25th Aug '94

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate to the degree of Ph.D.

Signed ..... Date .....

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Abstract

This work details two investigations into image capture, taken from the fields of X-ray and laser research, and also details two scanning systems: a wire surface generator and a video security device.

Firstly a camera system is described that can display images, digitize them and provide real time false shading. This camera is shown to have a linear intensity response and to have a maximum saturation level below the digitizing range. Some example outputs are then illustrated.

The ability to irradiate CCDs with direct X-ray radiation is also investigated. A camera is developed that vertically integrates such images and is shown to give an increase in the processing speed of existing equipment and to reduce experiment times by a factor of 388.

Taking this idea further, a fast one dimensional camera is developed. This camera couples laser pulses onto a CCD via a fibre optic faceplate and a 25μm slit. Unusual scanning techniques are used to achieve image storage within the sensor itself and a method for correcting dark current and other errors is proposed.

Next a mechanism for displaying wire surface representations of $(x, y, intensity)$ images is investigated. Results obtained from real time, hidden line removing hardware are illustrated, along with improved algorithms for shaded surface generation.

This is then developed into a security device protecting VDUs from radio based surveillance. This is achieved by randomizing the display order of raster lines along with a hardware solution for random sequence generation.

Finally the generation of Uniformly distributed random numbers is achieved by processing readings from a digitized, Normally distributed voltage source. The effects of this processing are investigated and an analysis of the underlying theory is used to determine an optimal setting for the gain stage.

Contents

List of Figures

Chapter 1

Introduction and Development Hardware

1.1 Introduction

There are many existing designs of CCD cameras ranging from those available in the high street shops up to those used for complex scientific laboratory equipment. Their applications now range from colour auto-focus household cameras up to large format dental X-ray systems.

One noticeable path in the development of CCD equipment has been that the format of the devices has been dictated, not so much by the application, but by the existing technology to which the CCDs have been applied. Hence the normal format of an inexpensive CCD is still a 288 by 388 pixel framing device suitable for TV applications. It has only been in the last few years that CCDs have been made for specific applications, and hence most applications of such devices concentrate on generating normal area pictures, usually for viewing via a display system.

This research aims to look at two CCD based instruments that do not follow the normal area image gathering format. In particular, scientific use of these devices will be looked at and the hidden problems associated with turning photon densities into digital signals.

Most CCD cameras used in laboratory instruments have been developed from a normal format camera by adding frame grabbing facilities and digitization units. These systems, rather than addressing the problems at their source, have usually been 'add on' parts, a CCD camera providing an analog signal which is converted into a PAL (or similar) format video signal. This is then passed into the frame grabber where

it is converted back into the analog signal and then digitized to generate a computer accessible image.

This has meant that, though many 16 bit digitizers are available, the signal to noise ratios of the cameras have in fact been far poorer than would be required to make use of this resolution and hence unrealistic claims have sometimes been made.

Thermally induced charge build up (dark current) within a CCD is another major cause of error that is rarely considered. Many applications take accurate measurements of signal strength but do not correct for the dark current distortions that have been included in such measurements.

Display hardware designs have also remained very static, usually following the normal interlaced raster scanning methods used for conventional TV and video monitor technologies. There has been little research into non-standard displays but, as will be seen, there are many possible ways of driving displays and these can illustrate significant advantages to the experimental analysis of image data.

Some work has been done into scanning systems that use vector rather than pixel based imaging but these devices have tended to be used in popular arcade games and some military applications rather than be put to use for general research purposes. There are many advantages to be gained from novel scanning of display hardware, including surveillance protection and reducing observed flicker. Most research in this field has concentrated on software solutions to the image rendering problems and only the BBC seem to have done significant work into the possibility of hardware modification or designs [1, 2].

Chapter 2 will look at a method of scanning a CCD that allows it to be used as a one dimensional sensor but that makes use of the area sensing pixels to increase the device's saturation level for X-ray photons. This looks primarily at novel frame grabbing hardware that utilizes a normal CCD signal in a non standard way. Particular attention is paid to an application of such a device to X-ray sensing, an area in which the CCD is finding increased popularity within the scientific community.

Chapter 3 will look at a further example of using a CCD as a one dimensional camera, in this case acting as a fast framing camera. This extends the ideas used in chapter 2 to use the actual CCD silicon as both the one dimensional sensor and the image store. This chapter also goes into detailed analysis of the effects of dark current

and digitization on camera images. Algorithms are developed to correct for the build up of dark current distortions both within the amplification stages and the digitization stages, and also to analyse the actual surface dimensions of the CCD.

Chapter 4 will look at a method of representing (x, y,*intensity*) signals, such as those coming from a normal CCD camera, in the form of a 3D surface. The need for such displays is becoming increasingly obvious and, with the advent of framing cameras which look at the time progression of (x, y,*intensity*) images, their use will become more common particularly as some interesting details can be seen from surface representations. The development of hardware and software to generate such images, quickly, from a CCD camera will also be described.

Chapter 5 goes on to look at a non raster scanning technique that allows video protection by scrambling the scan order of the output lines. A development circuit implementing a patent specification developed in the University of St. Andrews by Dr. P. Hirst and Prof. A. Maitland [84] will be looked at along with enhancements to the original proposal and an analysis of the waveforms required to generate an electromagnetically scanned output.

This will be taken a stage further in chapter 6 where a theoretical approach to the generation of Uniform random numbers from Normally distributed voltages will be undertaken. In particular the effects of digitizing continuous quantities will be investigated and the subsequent effect this has on the use of standard, statistical, conversion equations. An optimum choice of resolution for such conversions will also be examined.

Finally appendix A provides a brief analysis of existing CCD technology and covers some of the conventional driving electronics. The problems associated with both CCD input driving and output processing will be considered and some of the existing solutions to these problems will be described.

Firstly however, an analysis of the hardware developed to test out ideas and act as a prototype platform for experimental circuitry will be given before moving on to the developmental research work.

1.2 Development Hardware

1.2.1 The design of a general purpose development system

Before looking at the specific investigations carried out in this field it is necessary to look at the development hardware. Most of the original circuitry was developed around existing technology but a few points of interest emerged and these proved to be worth investigating in their own right.

In order to carry out any analysis of scanning techniques it was felt that a full and versatile development environment was needed. This was required to provide very flexible timing and pixel generation for both the driving of the CCD inputs and control of the display systems. This in turn requires a system that could provide variable frequency outputs for the CCD driving signals and any number of row or column clocks to drive both CCD and display inputs. The resulting hardware was based on the SCN2674 AVDC [3] (Advanced Video Display Controller). This component was designed to run PAL and NTSC displays but is very flexible and provides suitable outputs to drive camera and display controllers. Unfortunately it does not have a high enough clock frequency to provide full frame CCD control so a modified circuit had to be developed that enabled a faster clock to be sent to the CCD drivers.

In order to analyse the signals being input to the development hardware it was felt that an 'off the shelf' graphics and camera system would not be suitable. Manufacturers do not like to supply circuit diagrams and most of the internal functions of cameras and display systems are packaged within ASIC designs and hence are not adaptable to this type of development and in most cases the required signals are not even available.

Because of these limitations a simple graphics system was developed based on a graphics board previously designed in conjunction with Mr. I. Sheldon in 1988. The driving signals for this system were provided by a PC interface board that was developed by Dr. D.S.S. Robb for this purpose.

This board provided a Zilog Z80 microcomputer with an interface to a PC and a method of passing bytes between the processor and the PC. The Z80 provided direct access to the development hardware allowing the PC processor to carry out functions such as clock interfacing, without interrupting essential monitoring of signals such as the frame grabbing functions which will be described later.

This development system contained a block of memory that was driven by the SCN2674 and provided video images as a sequence of digital pixel intensities, each of which was an 8 bit parallel signal. The memory could either be filled from the PC, under control of a Pascal program, or from another digital source such as an 8 bit digitized CCD camera output. It could also be read by the PC at any time.

The 8 bit video output from this system was fed into the address inputs of another memory that allowed any mapping from the input 8 bit value to any 8 bit data output. The data inputs to this memory could be either from the RAM or from another 8 bit data source such as the digital camera signal to be described later.

The output of this mapping was then fed into a digital to analog converter and then on to a normal video monitor or, via a PAL converter, to a colour TV.

This flexibility results in the hardware being multi-purpose and allows it to act as a simple graphics board for the PC, as a frame grabber for a CCD camera or as a real time shade mapping for a live video signal. The main difference between this system and available frame grabbers is that the camera signal is never available as an analog video output but is digitized, in synchronization with the memory board, at its source. This allows the shade mapping of the camera signals to be achieved via the mapping RAM and also allows the digitizer to function precisely in synchronization with the display and timing control.

A camera system was also built that was driven by this hardware and provided a second source of the digital video signals. This was intended for use with the X-ray sensors described later but operated using standard format CCDs, in particular the EEV CCD02 series of sensors [4].

A full block diagram of this system is shown in figure 1.1. Most of this electronics follows simple design principles and details of the whole circuitry will not be gone into here. However, an overview of the different sections of this design is necessary to explain the origin of the driving and analysis signals described in later chapters. The following provides details and full circuit diagrams for each section of the development hardware.

1.2.2 Adapting the SCN2674 to high resolution applications

Figure 1.2 shows the circuit diagram of the timing and controller circuits. A three

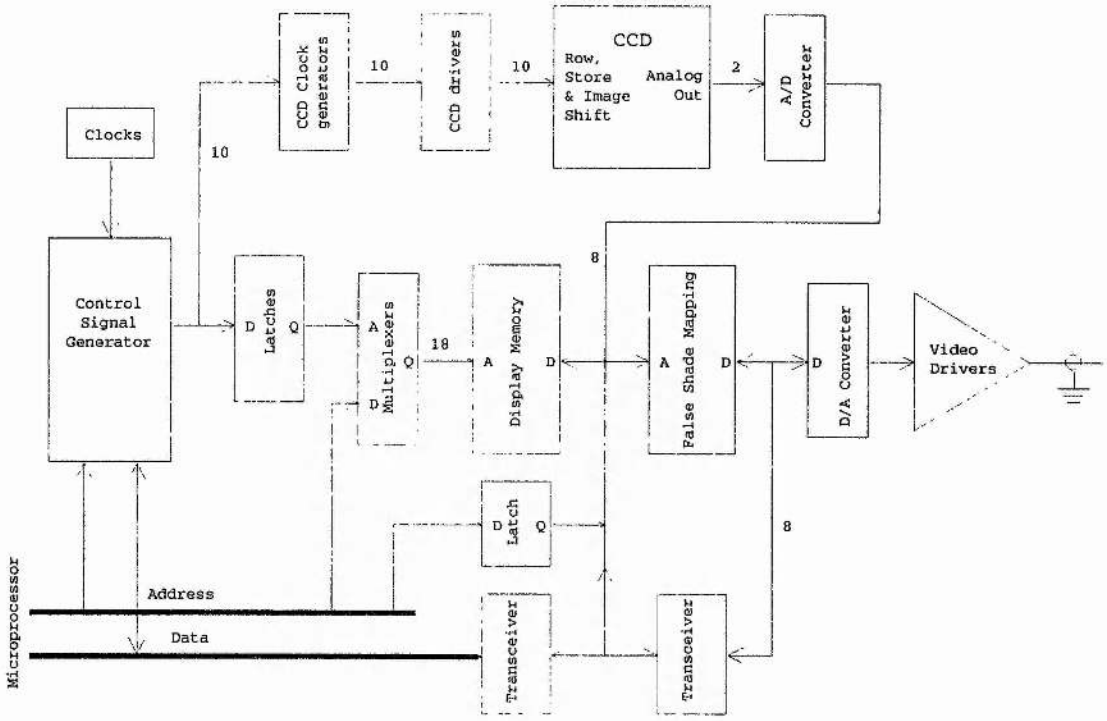


Figure 1.1: Block diagram of the graphics and display system.

phase clock is generated by a divide by three circuit and is then fed into a divide by four circuit. The outputs of this are fed direct to the lowest 2 bits of the addressing stages for the memory hardware (signals A_0 and A_1). For every 4 clocks into this stage a single clock is sent to the AVDC. This allows the AVDC to be driven from a clock frequency of around 2MHz (the maximum it can be driven at) provided a source crystal of 25MHz is used. The outputs of the AVDC then drive the address lines from A_2 to A_{16} and also the synchronization signals such as horizontal and vertical blanks.

As the effect of this is to output 4 pixels for every clock seen by the AVDC, all the settings within the AVDC will then be in multiples of 4 pixels rather than the single pixels it was designed for.

This circuit also provides stages to delay signals such as the horizontal blank and synchronization signals, so that they are correctly timed to drive the display hardware after the mapping and latching stages have delayed the digital video signal.

It should also be noted that the exact synchronization of these signals is achieved by latching them all simultaneously on a half clock from the original divider stages.

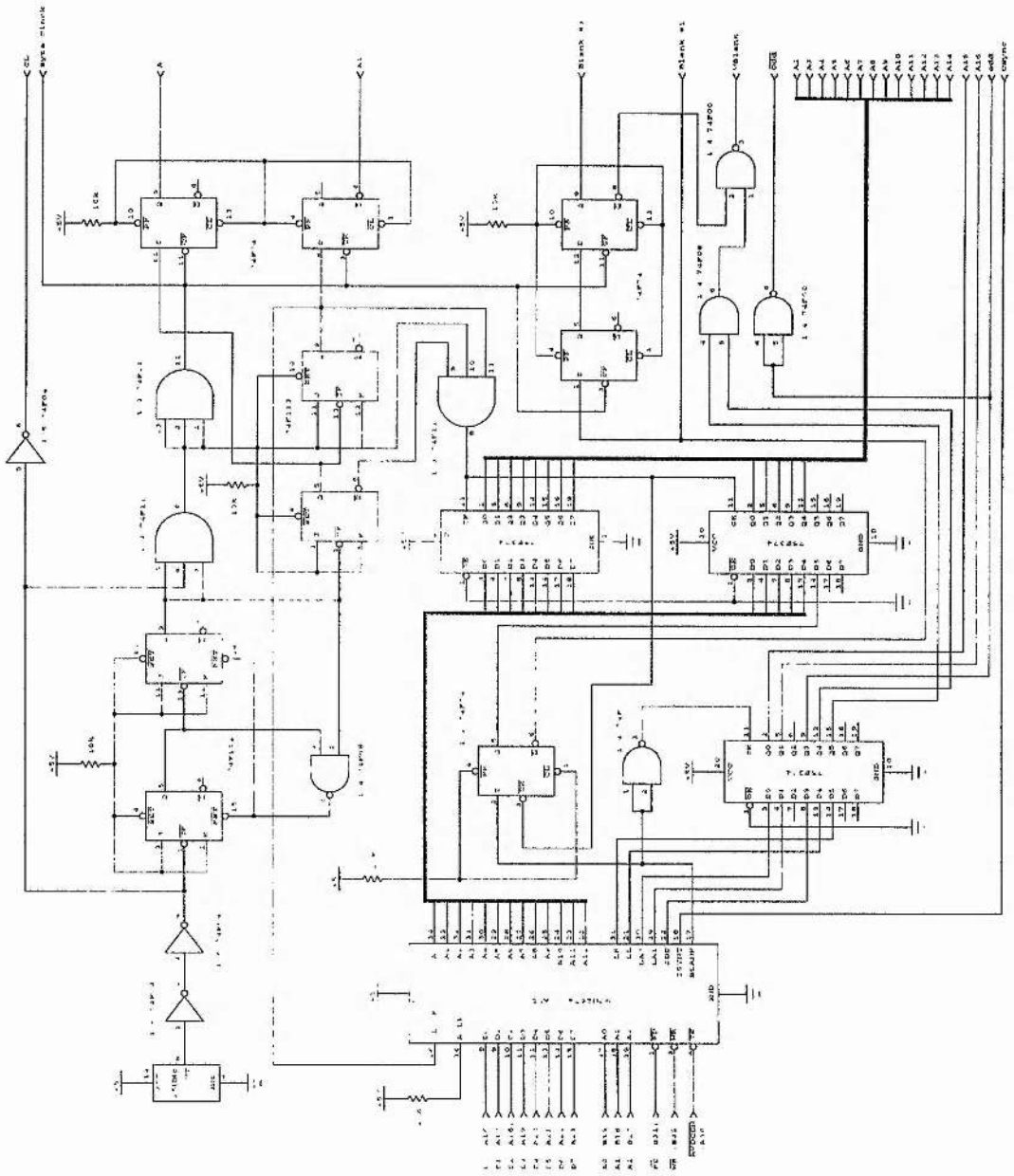


Figure 1.2: Circuit diagram of the timing and video controller.

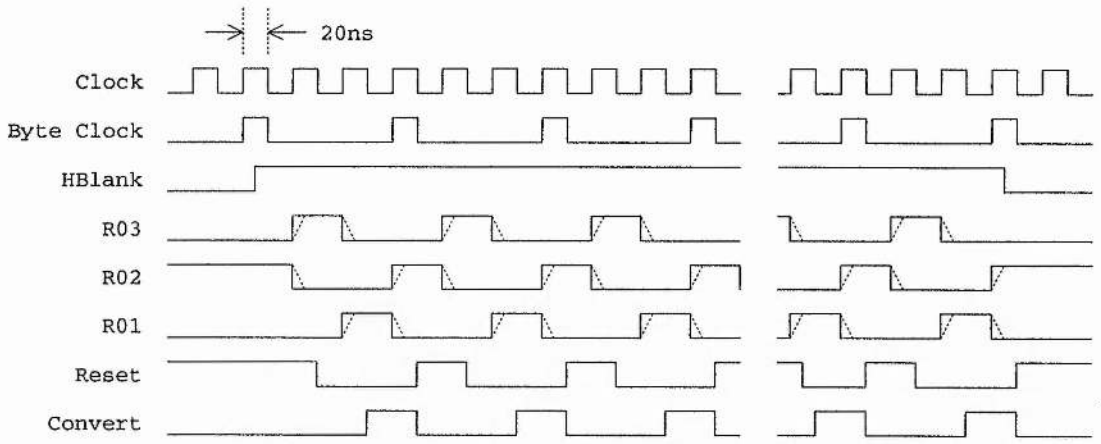


Figure 1.3: The timing of the video driver stages.

Figure 1.3 shows the timing of the output signals and illustrates how the output timings relate to each other. The slopes on the CCD drive signals are also shown.

An enable line from the PC was provided in order to access the AVDC and program in the required parameters for the number of characters per row, number of lines per screen, and constants such as equalization, blank period, synchronization times and front and back porches.

1.2.3 Frame grabber and display board

All the RAM addresses are multiplexed with the PC controller signals and are then fed to the four 32k RAM chips. The enable line from the PC is used to feed the multiplexers, ensuring that the PC has priority over the video signal. The output stages are latched so that, when the PC accesses the RAM, the video output stays constant at its last value rather than reverting to a random value or to zero. This is not the normal method of prioritizing a display output, as the user of a graphics systems does not want to have the image interrupted. However the purpose of this hardware is to provide an analysis tool and this was felt to be a suitable compromise.

The data outputs from this memory are then fed into the address input of the mapping RAM chip. Another enable line allows some of the PC controller address lines to access the address bits of this RAM and hence, along with the multiplexed data lines, allows the PC to put the mapping data into the RAM. A two stage setup was used that

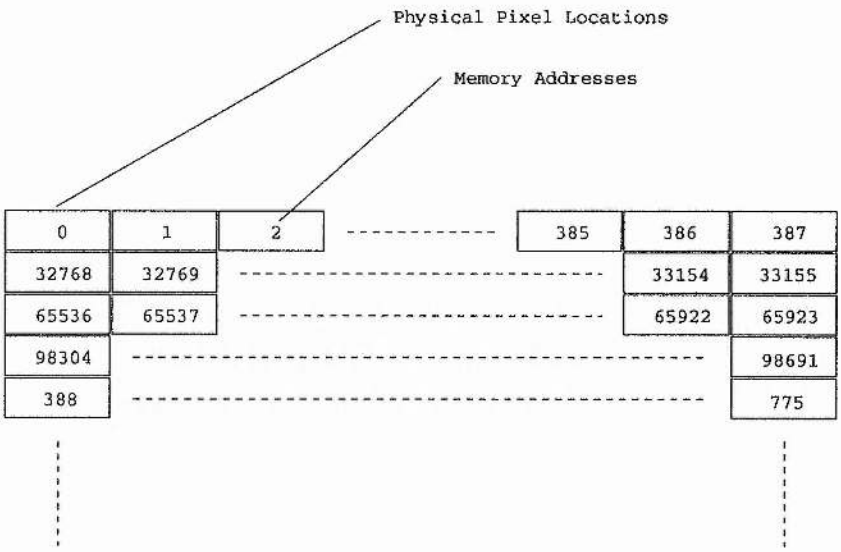


Figure 1.4: The mapping from pixel address to physical location.

allowed a different mapping to be enabled when either the memory, or the live video signal are being fed through this mapping. This is a hardwired line supplying the upper address bit of the memory. These signals were also multiplexed with the PC addresses.

1.2.4 Memory and pixel mappings

The memory addresses were arranged in such a way as to minimize possible interference from the decoders. The order of the address outputs from the AVDC is sequential, taking no account of the number of rows in the image, since the AVDC outputs addresses for characters not for pixels. Because of this it also provides 4 row selection signals to select which row of the character is being accessed. Two of these row lines are used to select which RAM chip is being accessed. These are fed into the highest address bits and hence into the chip selection circuitry. In this way the memory chips are interleaved, one for the first row, the next for row 2 and so on.

Internally the AVDC registers were then set to output 4 rows per character, 97 characters per row (one quarter of the number of pixels) and 72 rows.

Figure 1.4 shows how the addressing corresponds to the pixels seen on the display output.

Chapter 4 makes use of this feature to access only one quarter of the image by using

only one memory chip as a source of data and hence accessing only every fourth line of the image.

Initially the hardware was set up to use 388 pixels per row and 288 rows. This corresponds to a normal CCD camera system. Alterations to this were made when needed by the investigations covered in later chapters. The initial settings were for a non-interlaced display system. This reduced the resolution of the display but enabled the camera to be driven in such a way that the images were displayed with a one to one correspondence between camera pixels and memory locations. This is not normally the case with video digitizers that take composite signals. This system provides easy analysis of image coordinates and also reduces the flicker of the image.

Inserted figure I (see back cover) shows the circuit of the memory and colour mapping stages.

1.2.5 Logic control of the memory circuits

Another circuit added to the controller was one which produced a logical mapping controlled by lines from the PC and the development hardware. Four inputs were taken from the PC card, combined with the ODD frame output from the AVDC and then decoded to provide read, write and enable signals for the CCD data. The ODD frame signal is usually used for interlace purposes but in this case it serves as a line that changes state with alternating frames.

The three outputs of this circuit allowed a number of functions to be performed including writing of the CCD image into the frame storage memory (frame grabbing), output of just the CCD live signal, output of the stored frame and a mode that alternates between the frame image and the live image.

This latter allowed a picture to be overlaid with the live image. This meant that pixels could be set in the frame memory and overlaid very precisely onto the live image and hence allow very accurate measurement of features within this image.

The circuit of this section is shown in figure 1.5.

The frame grabbing function is achieved by the Z80 processor accessing the AVDC registers and waiting for a vertical blank. When this is detected these control lines are set to pass data from the camera into the frame storage. The AVDC is then monitored until the frame has completed and the vertical blank is active again. The lines are then

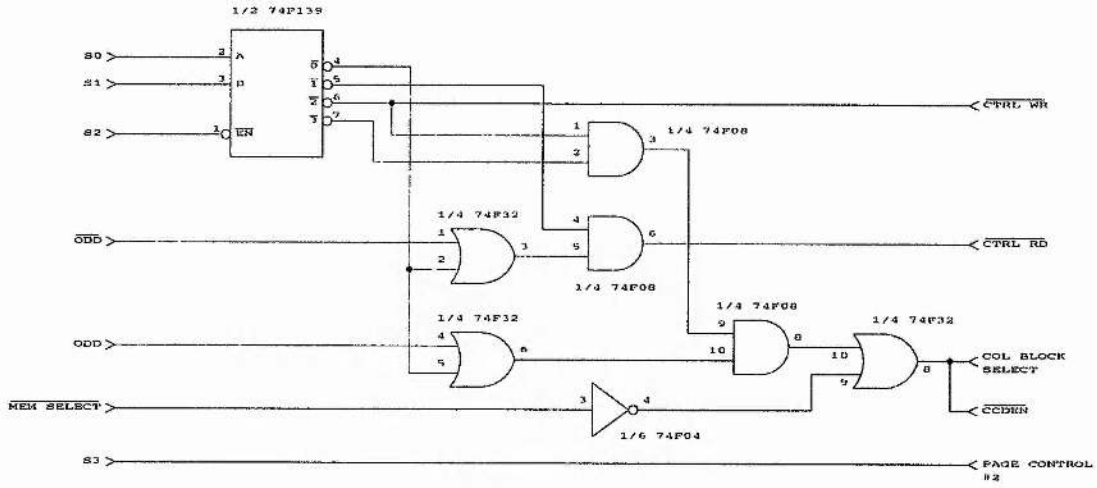


Figure 1.5: Logic control of the stored and live images.

Device	Speed	Resolution	Price
SP9770	5ns	10 Bits	£80·00
SP9768	12ns	8 Bits	£54·00
PNA7518P	33ns	8 Bits	£10·46

Table 1.1: Video rate digital to analog converters.

reset to complete the storage operation.

1.2.6 Video driving stages

In order to view the results a simple video output stage was added that provided conversion from the digital video signal to a normal composite synchronization signal for a standard monitor.

The output of the final latches of the memory circuit is fed into a high speed (30MHz) D/A converter. Table 1.1 shows the devices considered for this conversion stage. The PNA 7518 [5] was chosen as it is very stable and flexible as to the output voltages it can operate over. In particular it takes two input voltage levels corresponding to the upper and lower limits of the conversion range and these were set to correspond to the saturation levels for black and white. The output is then combined with the synchronization signal from the AVDC via a summing video op-amp which in

this case is an LM733 [6]. Again, this part was chosen from a selection of available products including the MPOP02, NE5534, NE5517, LM318, NE592 and the NE5205.

Finally the composite signal is sent through a 50Ω resistor required for impedance matching to the monitor. A resistor was added between the synchronization signal and ground in order to provide synchronization level adjusting. This allowed the signal contrast to be extended to match the type of monitor used.

Because of the sensitivity of the supply inputs to the D/A converter, a low noise, dual amplifier package was required. It was decided to use a RC5532A amplifier for this stage and the constant voltage feed was provided by an MP5010HN 1·22V bandgap device that provides a very stable output over a wide temperature range.

Figure 1.6 shows the full circuit of this stage.

Development with Mr. Sheldon added a PAL encoder and UHF unit to this in order to provide 2 or 3 bits per colour plane. This allowed false colours rather than false shades to be generated in the mapping RAM.

1.2.7 Power supplies

All the power supplies built both for this development hardware and the circuits developed in later chapters are based on a simple linear power supply design. The supply for this development electronics is shown in figure 1.7.

The analog and digital grounds were kept separate and only connected at the CCD development boards in order to ensure that a good star earth point existed.

The CCD bias levels are very critical and the operation is very dependent on the relative voltage levels. Because of this the driver voltages are set via an LM317L (low power) adjustable voltage regulator. The larger current requirements of the driver stages are provided from a higher power version of the same device (LM317T) and the normal 5V and -5V levels are generated from high power fixed voltage units that could easily supply the large digital section of the development hardware.

In order to minimize mains hum pick up from the power supplies, these voltages were generated remotely from the camera head and passed through to the sensitive electronics via shielded multi-core cables.

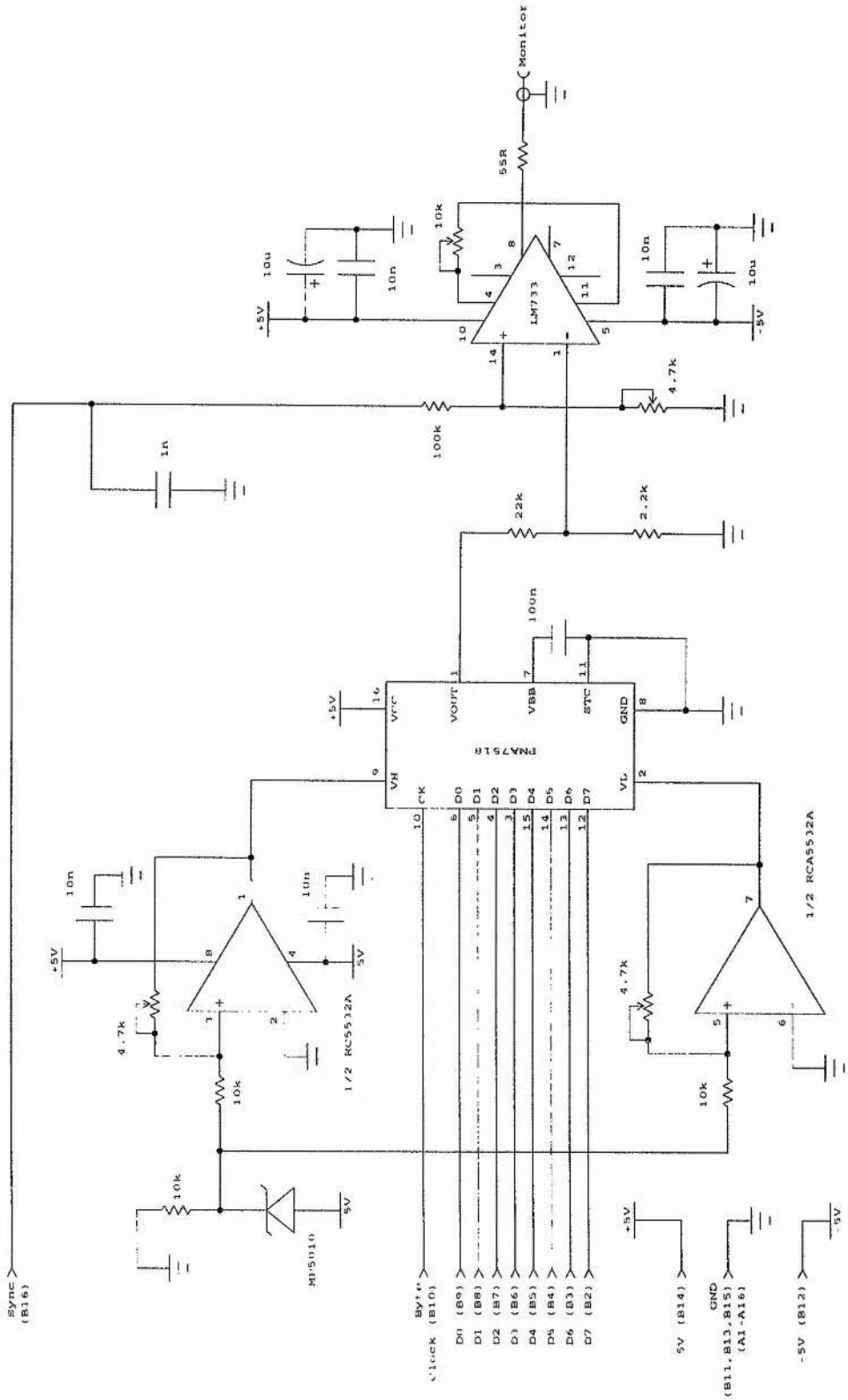


Figure 1.6: Circuit of the video driver stage.

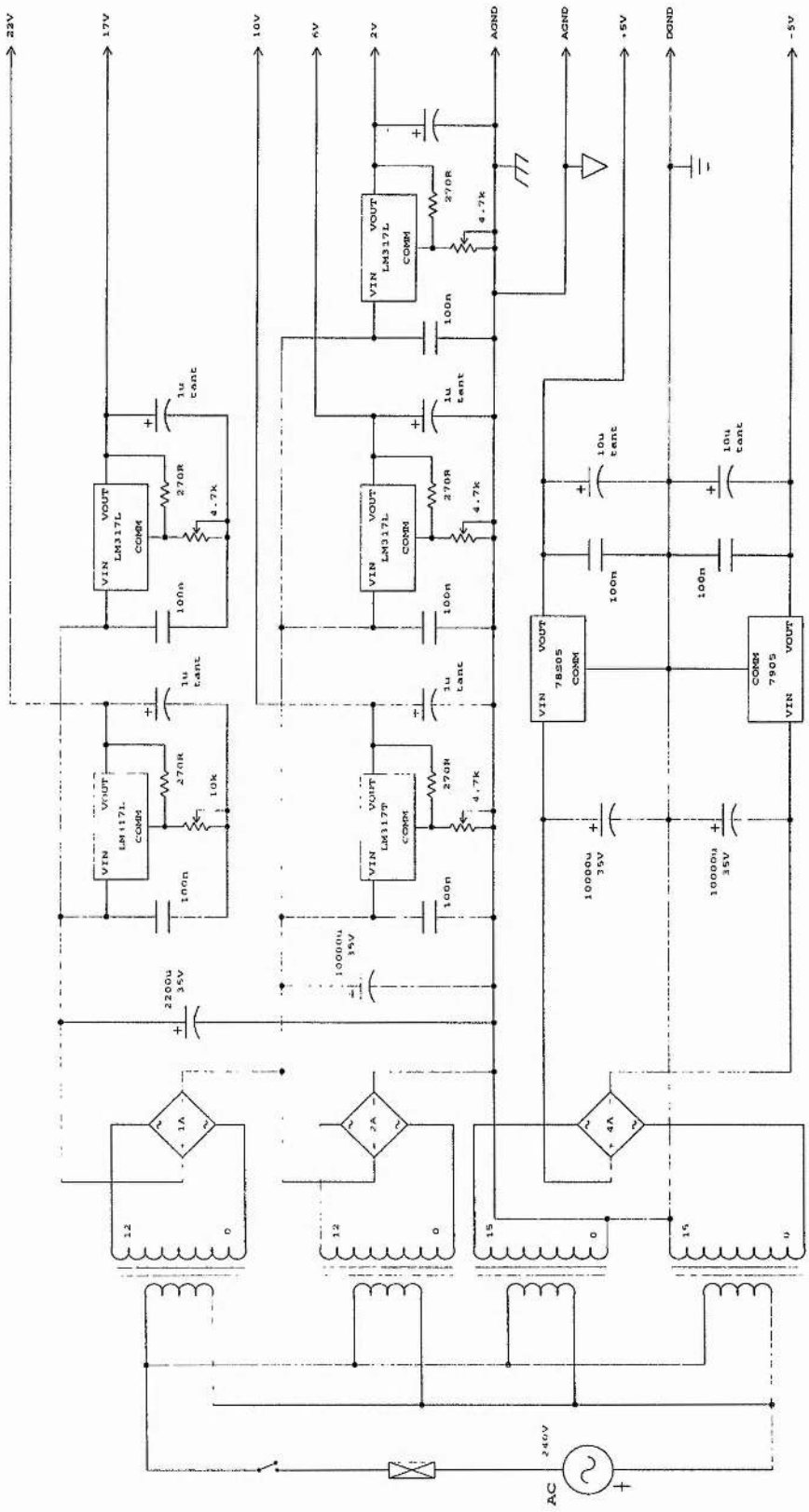


Figure 1.7: Power supply design for the development hardware.

1.3 A Digital Output Video Camera

1.3.1 The requirements of the camera system

In order to assess the image collection techniques described later a simple camera system was developed. This had to be able to interface to the synchronous digital signals and also provide digital output that the display circuitry could access.

The camera was mounted remotely to the display and its associated power supply circuits and hence needed to be driven from the minimum of signals. Because of the need to pass digital data back to the display hardware it was decided that ribbon cable would be used but every alternate strand was grounded thus forming a simple screen in a similar manner to the twisted pair screening method.

1.3.2 Digital driving circuitry

As described in the previous section, the AVDC enabled a completely variable set of timings to be used for the display hardware. The circuitry controlling the timing allowed any individual pixel to be accompanied by a full three phase clock to the CCD row shift inputs. The camera driving was achieved by a set of digital logic gates that took only the vertical blank, horizontal blank and the CCD phase clock (the 25MHz signal). These were combined into the horizontal clocking signals and also into the vertical shifts required to clock the CCD image down into the row readout register. The vertical blank, in particular, is used to trigger the image shifting from the exposed area into the image storage section of the CCD.

Inserted figure II shows the circuit diagram of the digital stages. It is based around three 74LS194 shift registers, one for each output line, and two 74HC4040 counters [7] for the frame shifting section. The use of AND gates on the inputs improved the signal rise and fall times after they had been transmitted down the connecting cables.

1.3.3 Analog driving circuitry

The conversion from digital (5V) to CCD driving signal is done using the DS0026 clock driver chips [8]. This choice will be detailed further in chapter 3. There are a number of possible driver chips, including the Sony BX1432A which is a video processing module, the ICL7667 from Intersil and the EEV ESB3650 series. The EEV components were of

Device	Speed	Resolution	Price
MP7683XKN	5MHz	8 Bits	£45·00
MP7684KN	10MHz	8 Bits	£57·28
CA3318C	20MHz	8 Bits	£72·00
ADC300	20MHz	8 Bits	£33·26
ADC301	3MHz	8 Bits	NA
ADC304	20MHz	8 Bits	£10·00
ADC208	15MHz	8 Bits	NA

Table 1.2: Video rate analog to digital converters.

particular interest as they form an entire camera controller which includes timing pulse shaping [9] and video processing [10]. Unfortunately too few of the required outputs are available to outside electronics and hence they could not be used for this project.

A stabilized power supply to the DS0026 drivers enabled good edges to be imposed on the output waveforms with the slopes for rising and falling edges being created by RC networks that act as filters to round off the edges and produce the required critical slopes.

The input capacitors allow for an increase in the rise times for the faster clocks and the use of smaller output resistors improves the filters for the higher frequency pulse shaping.

1.3.4 Initial amplification hardware

For this design a high frequency amplification and digitization stage was required. The output of the CCD had to be sampled and then held for input into an analog to digital conversion stage.

It was decided to use an ECL sample and hold unit and then feed the output from this into a digitizer. Three possible components were considered the SHM360 or SHM361 from Datel and the HUMC (25ns) or UH3 (30ns) parts both available from a number of alternative sources. The component chosen was the SHM361 from Datel [11] and the clocks were driven through an MC10124 TTL to ECL converter [12] and then back to TTL via an MC10125. This unit had a clock output and this was converted back to TTL levels and input to an ADC304 converter [13]. Table 1.2 shows the selection of alternative converters available at the time.

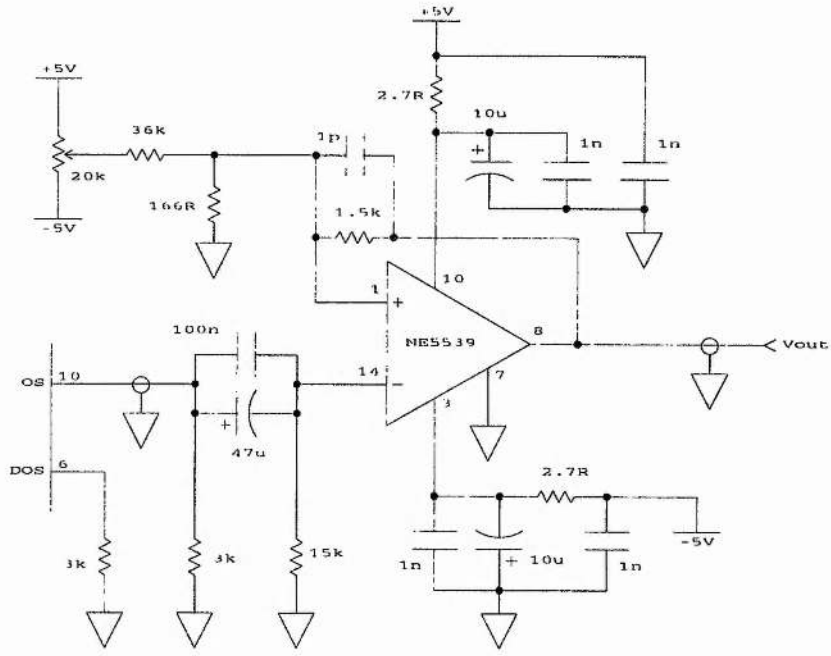


Figure 1.8: A higher frequency amplification stage.

The differential amplification stage was provided by an LM733 amplifier, a common component for use with video amplification stages. The circuit diagram for this version of the digitizer is shown in inserted figure III.

1.3.5 A higher frequency amplifier stage

Analysis of the output of this circuit showed that the capacitive effects reduced the bandwidth of the output signals and hence led to poor quality output from the camera. In order to improve this a higher frequency stage was built using an AD5532 video amplifier [14]. This is a very high bandwidth (4GHz for small signals) device that allowed the fine detail of the CCD output to be analysed. This was further improved by incorporating it into a dedicated PCB. This improved the noise characteristics of the circuit and its schematic is shown in figure 1.8.

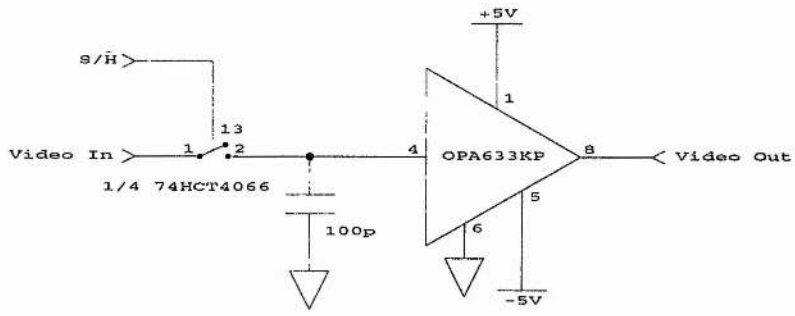


Figure 1.9: A discrete sample and hold circuit.

1.3.6 Discrete sample & hold

Problems with the sensitivity of the ECL devices led to the design of a discrete sample and hold circuit.

This circuit was based around an existing system designed by Mr. J. Wade at The University. The effect of this was not satisfactory however, as the noise added to the signal degraded the results more than the sampler improved the circuit and in addition the input impedance caused leakage of the signal. The appropriate element of this circuit is shown for reference in figure 1.9.

One important point that was observed here was that the CCD output signal was stable for longer than the conversion time of the digitizer. This meant that, provided a very stable output signal could be obtained, then the converter could be connected directly to the amplifier outputs and produce good results.

In order to achieve this two improvements were needed. Firstly a voltage reference was needed for the A/D converter as the now redundant SHM361 had provided this stable voltage source. Secondly the use of a multiple buffer stage was required to improve the signal amplification and allow direct connection of the converter.

1.3.7 Constant voltage source

A simple voltage reference, similar to that used to drive the video output, was suitable for the ADC304. To reduce noise, an OP27 low noise instrumentation amplifier was used and the MP5010 1·22V bandgap source provided the stable input. The gain was

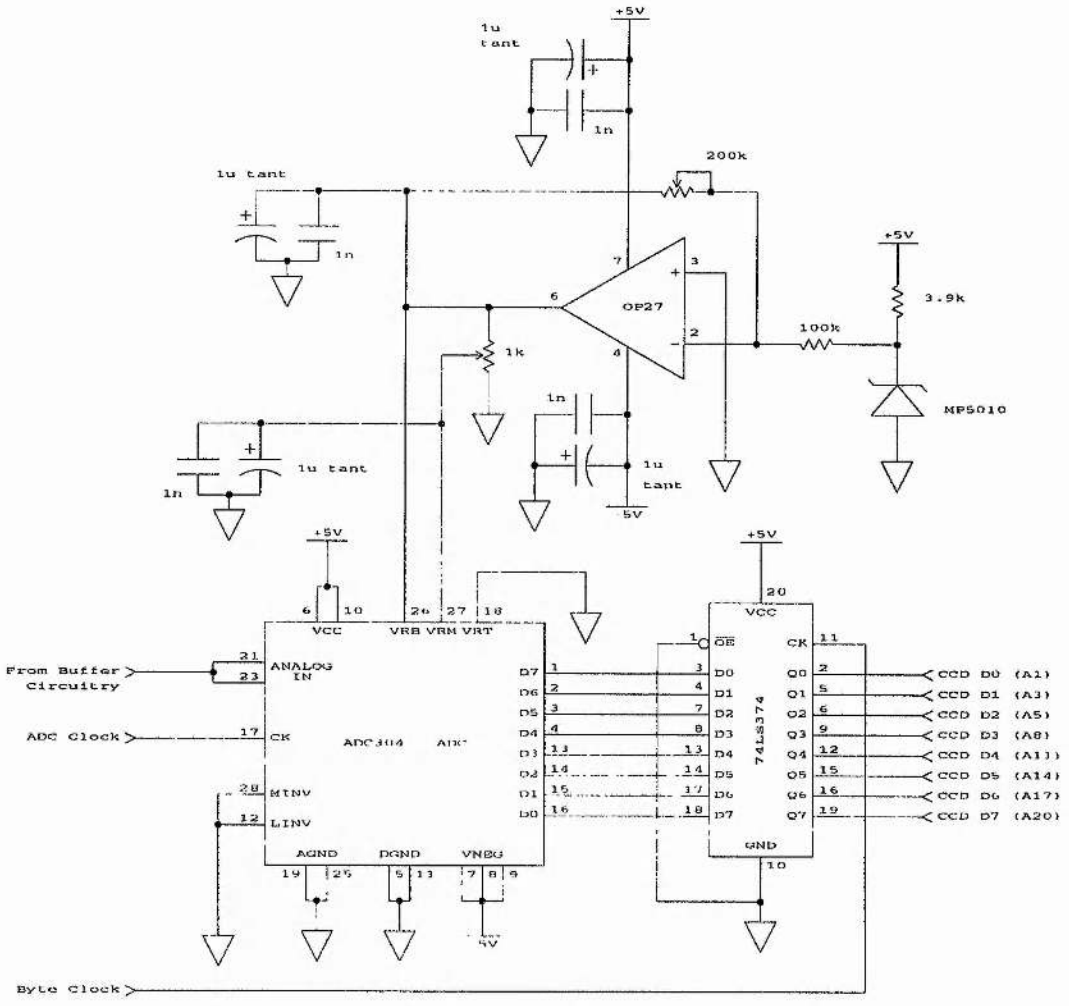


Figure 1.10: The digitizing part of the camera circuit.

adjustable via a potentiometer and provided fine adjustment of the conversion levels. The entire circuit was then designed onto a PCB to minimize noise.

Figure 1.10 shows the final circuit of the conversion stages. As can be seen from this diagram a latch was added to provide stable output signals and to provide drive for the cable lengths back to the display hardware (about 0·5m).

1.3.8 Buffered high frequency amplifier stage

A number of buffer stages were considered. The buffers had to be suitable for the high frequencies involved (50MHz or so) without causing noticeable distortion. They

Device	Bandwidth		Input Impedance	Supply	Manufacturer
	Full Power	Unity Gain			
Unity Gain Products					
EL2001	70MHz	-	8MΩ	±5V	Elantec
EL2002	180MHz	-	2MΩ	±18V	Elantec
HOS 50 [15]	100MHz	-	NA	±10V	National
Amplifiers					
OPA633	260MHz	40MHz	NA	±12V	Burr Brown
OPA620	100MHz	NA	NA	±5V	Burr Brown

Table 1.3: High Frequency buffers and amplifiers.

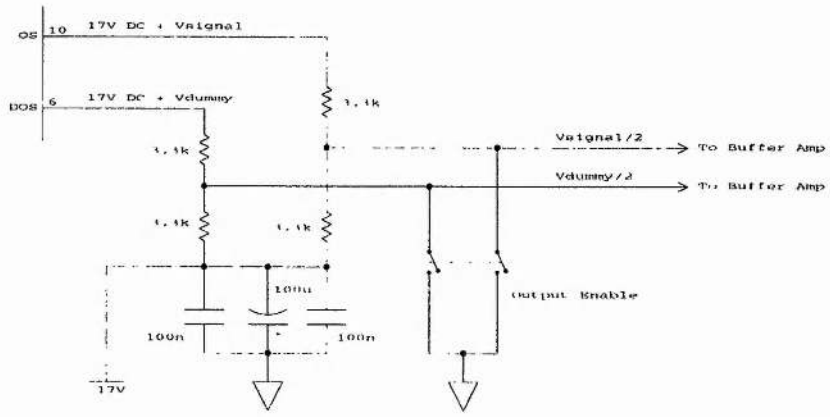


Figure 1.11: Removing the DC offset from the CCD outputs

also had to be able to take the signal levels of about 200mV. Table 1.3 shows the components considered.

The final choice was to use the EL2001 unity gain device as this easily met the required bandwidth and has good input impedance. The OPA620 amplifier [16] was then chosen to provide the gain stages.

1.3.9 DC removal by balancing with a negative supply

The problem of removing the DC offset from the CCD outputs was solved by using a negative supply of -15V and feeding the CCD output signals through two matched resistors in series, to this negative voltage level. This is shown in figure 1.11

The output DC signal is now reduced via the voltage divider, to a signal that can

be fed into the sensitive inputs of the buffers. This technique does not appear to be used in camera design, presumably because the presence of negative supplies requires considerable circuitry overheads.

The major advantage achieved by this is that there is no signal dependent distortion that is normally present when capacitive DC blocking is used. There is however a fixed distortion in the signal but this is a systematic, rather than random, error and is corrected for by the amplifier stages before passing into the digitizer.

The disadvantage of this system is that the required AC signal is halved by the voltage divider and hence the signal to noise ratio is halved. Provided very stable and smooth negative supplies are used, and they are very well decoupled at the source, then this is not a significant problem.

A DC offset was also added to the circuit to correct for imbalances in the amplifiers and offsets in the gain stages. Again this was also adjusted by a potentiometer for increased accuracy.

The final CCD decoupling and amplifier stages are shown in figure 1.12. Here the use of the buffers is illustrated, both on the CCD outputs before the differential amplifier, and on the output of the second stage amplification to drive the A/D converter.

The decoupling of the power supplies to both the CCD and the analog circuitry was found to be essential and tantalum bead capacitors were used to improve the smoothing effects for fast transients. A large (100μF) capacitor was used to smooth the negative voltage source obtained from a bench supply.

In order to protect the CCD outputs the circuit was powered up with the divider shorted to ground by a two pole switch. This was then switched out of the circuit when all the supplies were stable and the negative voltages could not provide damaging overload on the CCD FET output stages.

Two amplifier stages (a difference amplifier followed by a gain stage) were used to keep the bandwidth of the circuit high enough for the converter to have time to digitize the signal. As the converter was being used in its negative supply mode (converting from 0 to -2V), this second stage also acted as an inverter amplifier.

The whole buffer and amplifier stage was made as a double sided PCB using the component side as a large ground plane. This made a very significant improvement to the noise characteristics of the camera.

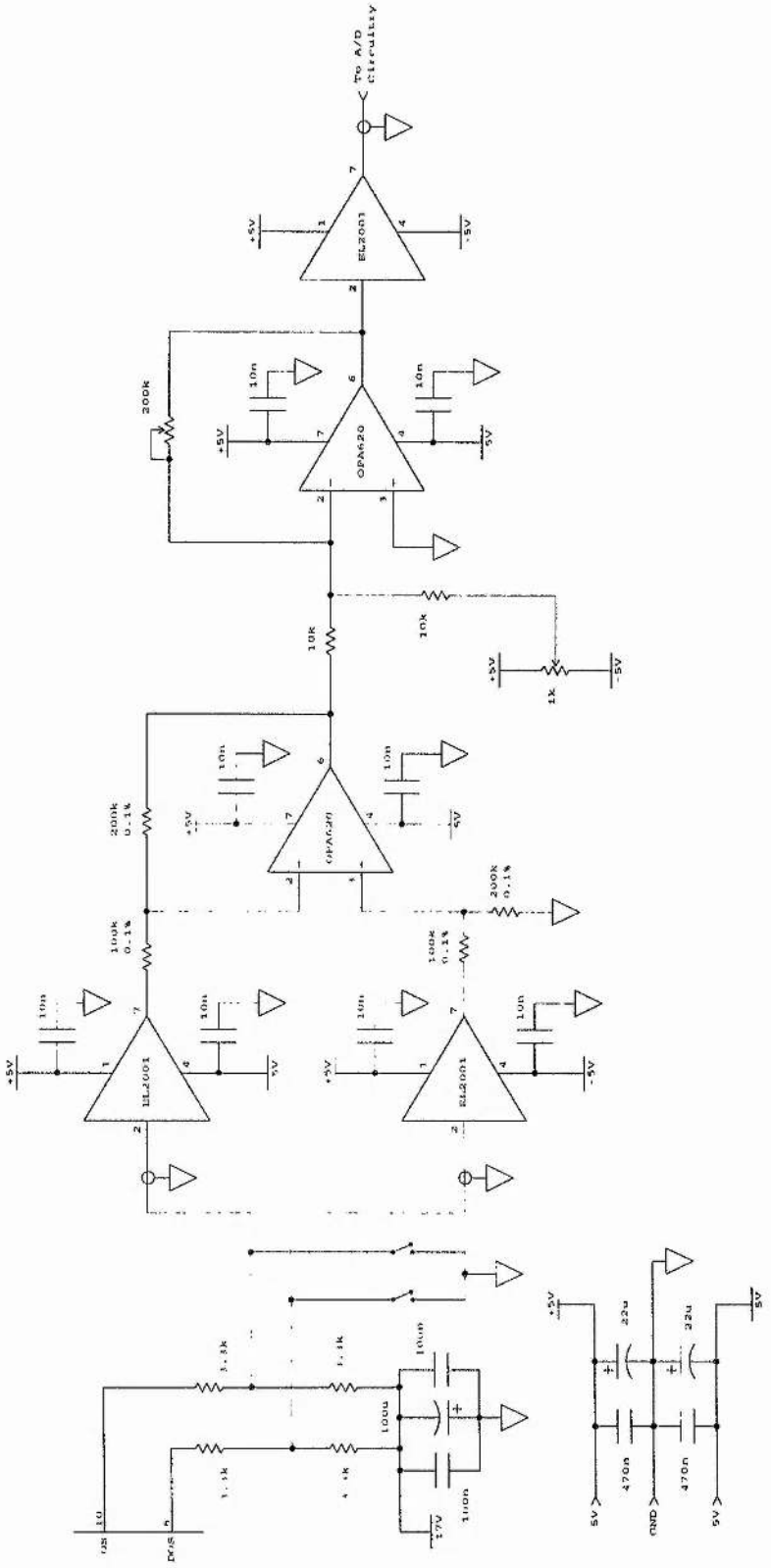


Figure 1.12: Buffer and amplifier stages for the CCD camera.

1.4 Analysis and Use of the Camera

1.4.1 The use of the shade mapping hardware

The advantage of false colour or false shade hardware is that it allows more detail to become visible to the unaided eye.

The controlling software written for the camera was fairly complex and allowed a large number of operations to be performed on the images. These included:

- downloading images to the PC for storage,
- loading of stored or created images from the PC to the display,
- switching the hardware between the different modes of display (live camera image, display memory image, frame grabbing etc),
- loading of a variety of colour maps.

All the pixels of the images are grabbed as bytes, 0 representing black and 255 representing white. The shade mapping allowed this to be changed to say, inverse (mapping 0 to 255, 1 to 254 and so on). Figure 1.13 shows an image frame grabbed from the camera and then displayed on the display hardware. This image was created by drilling a few holes into a brass plate which was then mounted in front of the CCD and illuminated with a diffuse light source. The source was adjusted so that the resultant image just reached saturation at its brightest point while still keeping the background at saturation black.

The edges can be enhanced by using a shade map that maps 0 to 0, 1 to 32, 2 to 64, 3 to 96 and so on up to 7 which maps to 224. The pattern then starts again with 8 mapping to 0 and so on. This produces the image shown in figure 1.14.

By mapping 0 to 0, 1 to 255, 2 to 0, 3 to 255 and so on, the image in figure 1.15 is obtained. This shows the full extent of the image but loses detail of the underlying structure that makes up the brighter areas.

A routine was also written that mapped all the shades to 0 except for 0 itself which was mapped to 255. This was displayed for a short time and then 0 was mapped to 0 and 1 was mapped to 255. This continued all the way through the shades and allowed the observer to see how the intensity areas changed throughout the image.

A similar technique allowed only every 10 shades to be turned on, having the effect of contouring the image.

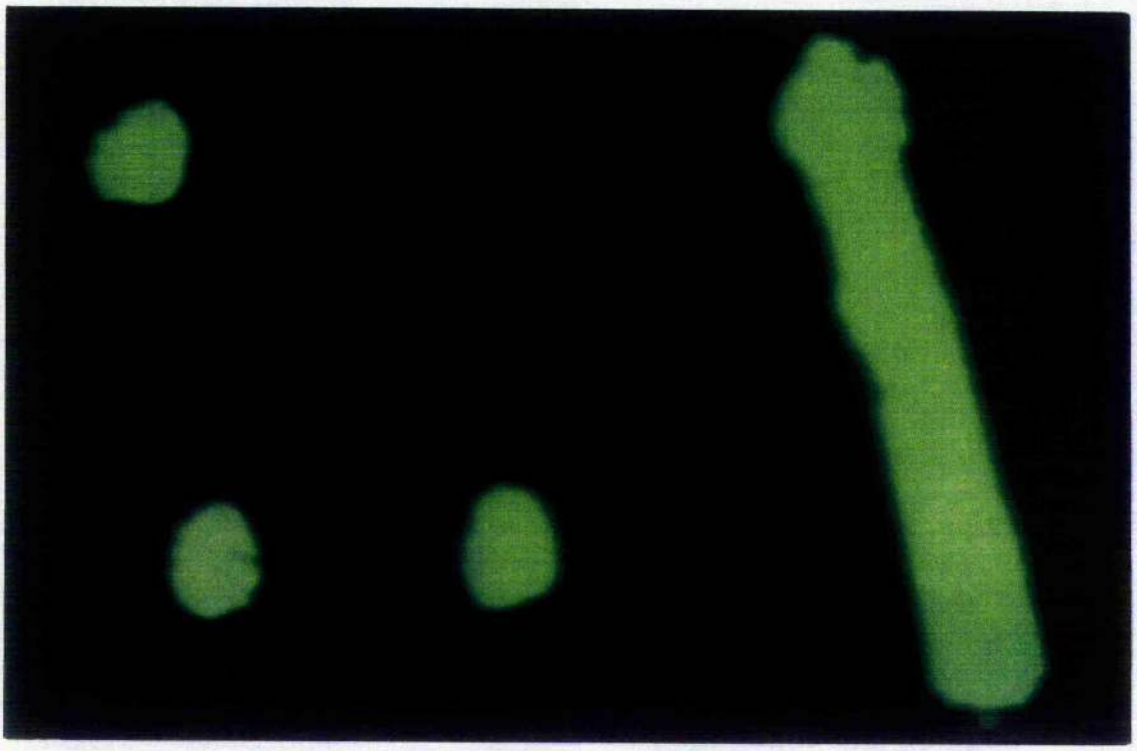


Figure 1.13: A simple frame grabbed image from the camera illustrating both saturated black and saturated white regions.

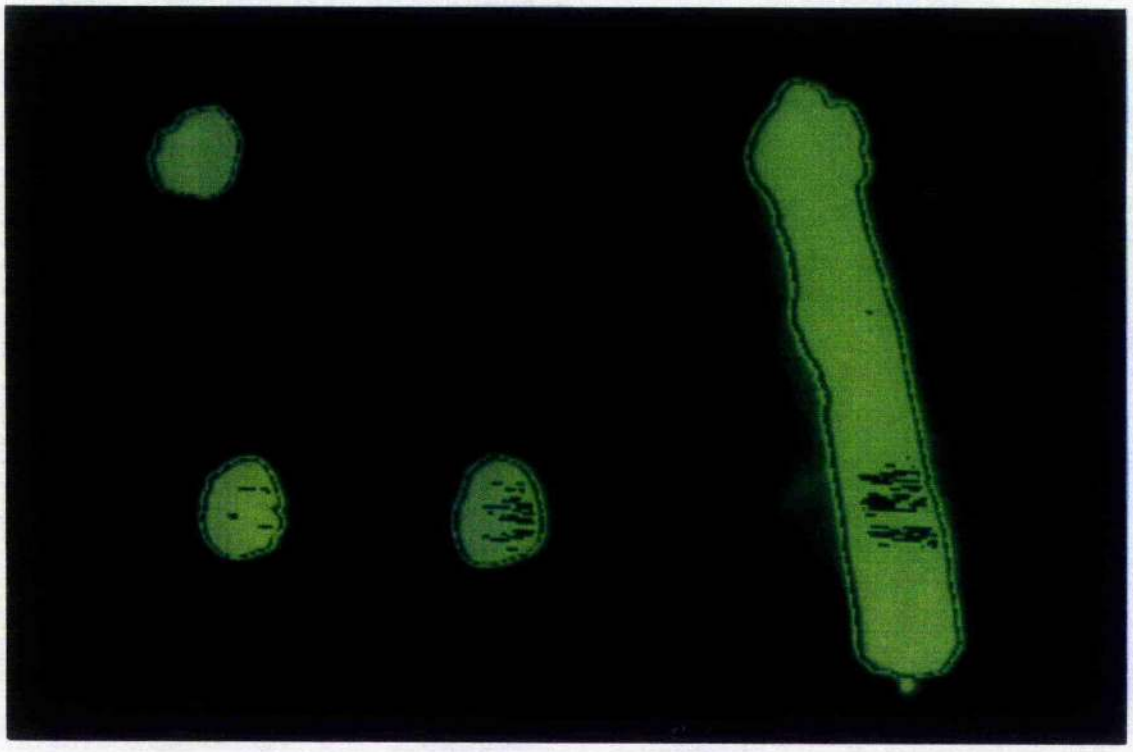


Figure 1.14: Enhancing the edges using the colour map .

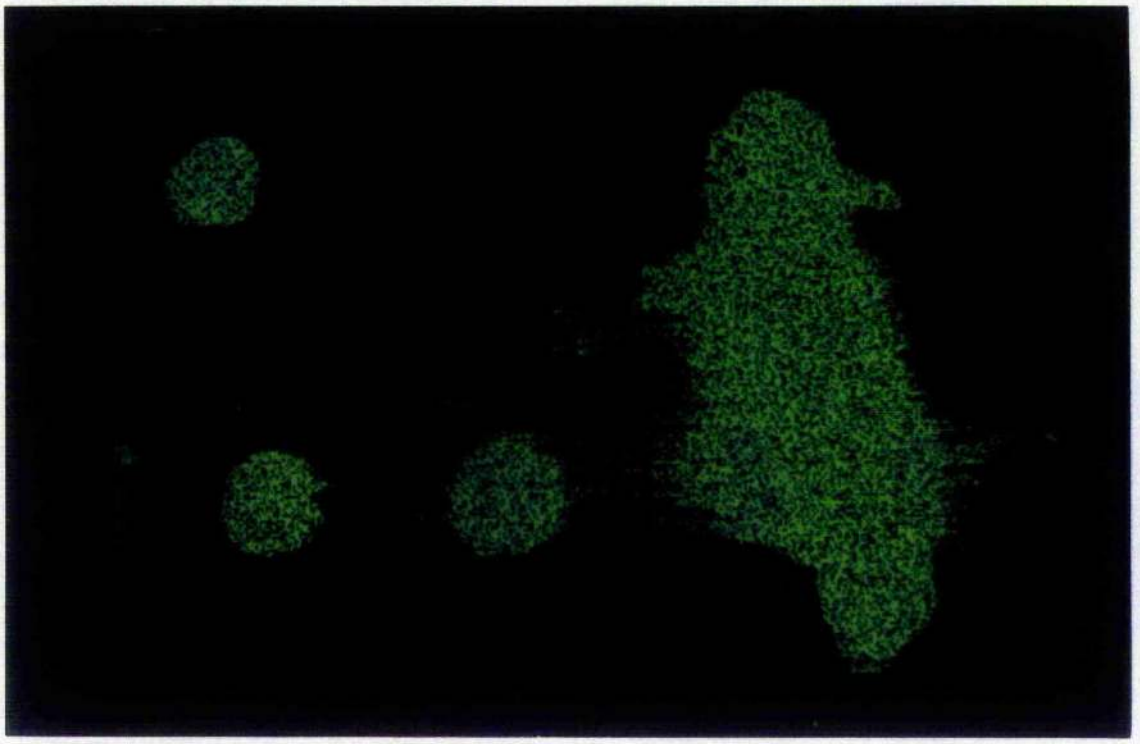


Figure 1.15: Using the colour map to show the full extent of the image.

As can be seen from these images the camera is quite stable and good results were obtained from the use of normal CCDs.

One other form of image enhancement is to display the image as a surface. Routines were written to do this via the camera and display hardware but a much more detailed investigation into this form of enhancement will be carried out in chapter 4.

1.4.2 Calculation of offset and noise for the camera

In order to evaluate the camera performance the image was first adjusted so that in total darkness a slight background 'haze' could be seen. This meant that the camera generated a clearly visible intensity floor, allowing more accurate analysis of the image noise.

The amplifiers were then set so as to not quite saturate the image for full brightness on the CCD. This meant that the full range could be properly analysed.

Once correctly set up a reference dark frame was taken that allowed software processing to find the offset level on the image by simply finding the mean pixel intensity. The mean was found to be 18·003 ADUs (analog to digital converter units).

It was also possible to find the noise levels in the camera by looking at the deviation from this mean value. From this program the variance was found to be 3·337 ADUs and hence the standard deviation was 1·827 ADUs.

1.4.3 Driver and analysis software

The controller software also provided the ability to compress the data downloaded from the image.

The algorithm used here was developed for smooth images such as those obtained from the X-ray experiments analysed in chapter 2. These images have the useful property that they are smoothly shaded in both the x and y axes.

The algorithm operated by finding the first row of data, run length encoding it and sending it to the output. For the subsequent rows, the differences between this row and the previous row is found, these differences are then run length encoded and finally Huffman coded. The resulting codes are then sent to the output.

The result of this method is to make use of both the horizontal and vertical regularity of the data to improve the encoding reduction. In some simulated images, results of up to 95% reduction were achieved.

Experiments with irregular images showed the weakness of this algorithm and results of only 5% or 6% reduction can be expected for discontinuous images.

Other routines to work out minimum and maximum values, number of occurrences of each value and so on were also written to aid in the development work and the analysis of resulting images.

Some test images were also created using shaded areas and lines. This enabled the vertical summer hardware in chapter 2 to be tested and also provided test images for the surface generators in chapter 4 and the scrambler detailed in chapter 5.

Chapter 2

Vertical Integration of X-Ray Photons Using a Charge Coupled Area Sensor

---

2.1 Introduction

2.1.1 Generic crystallographic structural analysis

One of the most interesting applications of CCDs is in the field of X-ray detection. As will be shown later the silicon provides a large gain to the X-ray signal and also has good quantum efficiency in the soft X-ray region.

Within the field of X-ray crystallography there are a number of applications for X-ray detection. Area detectors are most useful in applications where a large section of the reflected X-ray beam is to be analysed, but so far they have not found their way into applications where the experiment requires a very restricted analysis of the beam in only one dimension.

Intensity readings for large area experiments are usually done using two angles. Once an intensity reading has been taken the detector is rotated through a small angle ($\Delta\theta$) and a new reading is taken. Once a full range of intensity against detector angles has been measured, the crystal itself is moved through a small angle ($\Delta\omega$) and the angular movements of the detector are repeated. Computer settings and the use of stepper motors allows both of these angles to be accurately set up on a universal mount such as the CAD4 X-ray diffractometer universal mount.

For many applications, particularly with biological crystals, the angles scanned are quite large. These experiments look at the spatial distribution of Bragg reflections

caused by the crystal's atomic structure. The position in space of these spots, and their intensities, can be analysed to produce the atomic positions. The analysis requires vast amounts of computation and is usually done by computer. One example of this type of detector is the FAST[†] system used by Dr. Uli Arndt at MRC laboratories Cambridge [17]. Here a large area sensor is used enabling the detector to be fixed in one place and making use of the size of the detector to cover all positions in one setting. These experiments are slow, sometimes taking 8 hours. The software suite used at Cambridge is the MUDD system, written in FORTRAN and processing in this environment can take up to 24 hours to complete after the data collection has finished.

Some initial work into using CCD detectors for this application has also been done by the MRC research groups looking at both the distortion correction of normal, SIT tube, detection [18] and also the use of CCDs as X-ray TV detectors [19, 20].

This illustrates the first major problem of this type of detector. The time taken to collect the data means that biological crystals tend to denature producing radicals and gradually losing their structure. With the availability of very high intensity sources such as the CERN accelerator, it is now necessary to develop faster systems that can replace the SIT tubes used by these detectors.

2.1.2 The analysis of a single Bragg reflection

Apart from the large area experiments described above, there are some experiments that look at individual reflection spots, rather than the spot distribution. These attempt to map the intensity profile around a single spot and hence look at the electron distribution within a molecule. One such experiment is being pioneered by Prof. A. Mathieson of La Trobe University [21, 22].

Figure 2.1 shows the schematic of the experimental apparatus that is used for most crystallography detectors where such measurements of the Bragg reflections are required. The normal beam used is a $0{\cdot}5\text{mm}^2$ collimated X-ray source generated from $0{\cdot}7\mathring{A}$ Molybdenum or $1{\cdot}5\mathring{A}$ Copper emissions from a 1kW X-ray tube. This beam strikes a test crystal and the diffracted beam then goes on to strike a detector.

Traditionally, for small area analysis this detector is a NaI crystal (coated with aluminium to reduce reflections) that forms part of a scintillation counter. This produces

[†]Developed by Enraf-Nonius, Holland.

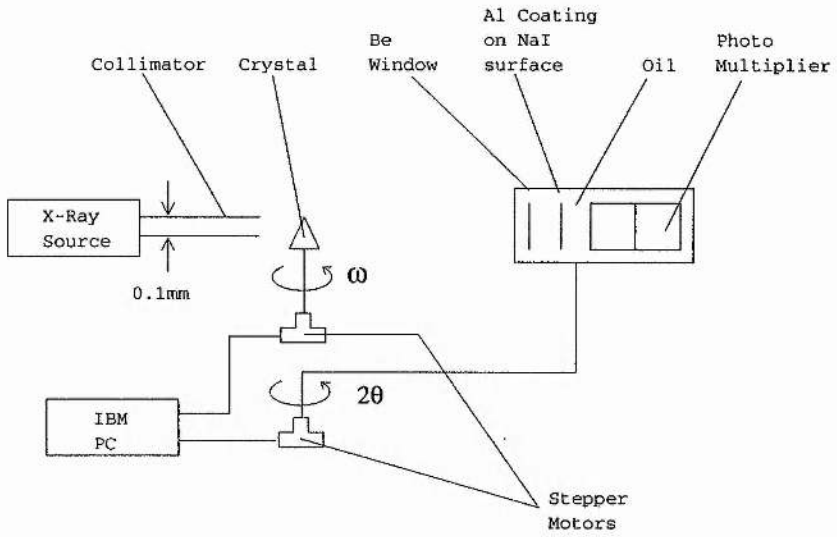


Figure 2.1: Apparatus used for single Bragg reflection analysis.

approximately 300 quanta of light per incident X-ray photon.

This light is then passed through oil to enhance the optical coupling into a photomultiplier. This in turn feeds a discriminator and is then sampled. The X-ray energy is detected by the number of photons the scintillation counter produces. The counter has an upper limit of 10000 counts per second before saturation and if this is exceeded then the further counts are lost. This is one of the limiting parts of such apparatus and defines the minimum time in which experiments can be done.

Initial work into using the FAST system for this type of experiment [23] showed that the device, though just able to gain the required resolution, took a very long time to collect sufficient data.

Another problem is that, though the incident X-ray photons cause a large gain to occur in the phosphor coating of the SIT tube, at the same time the directionality of the photons is lost. This means that the emitted optical photons come out of the phosphor at almost any angle and give rise to what is known as the point spread function (PSR) of the detector. Figure 2.2 shows this effect. There are many problems associated with this and it is the main cause of low resolution in many detectors. The 'tails' of the emitted Gaussian light cone can extend over a very wide area and in some cases, such as the FAST unit, these tails can be detected over the entire viewing frame.

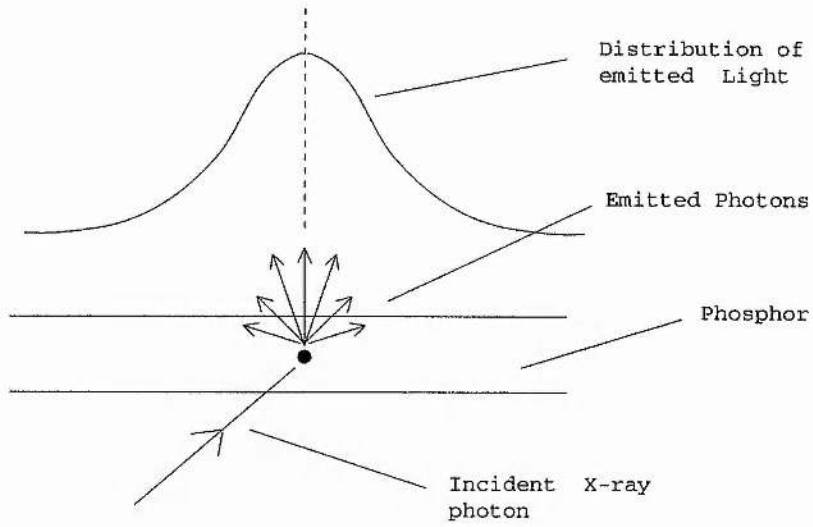


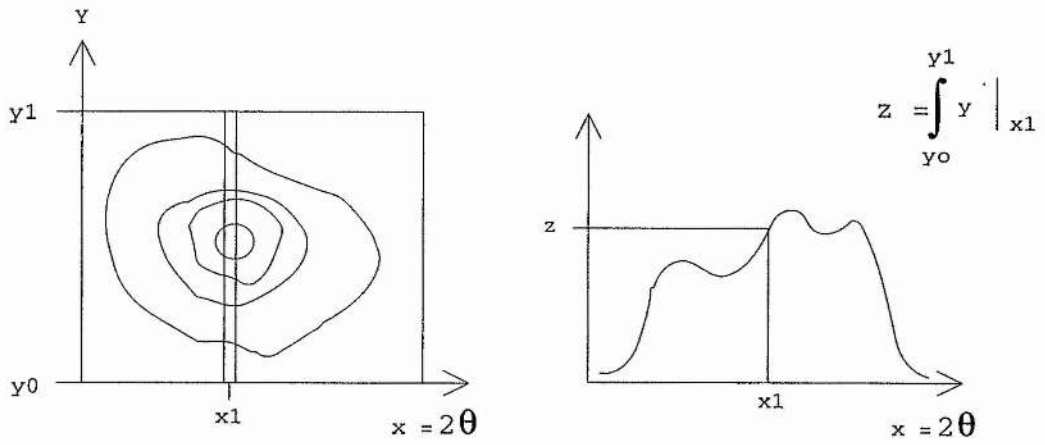
Figure 2.2: Point spread function in a phosphor.

The apparatus for single reflection analysis differs from other systems in that only a very small angular movement is used and hence only one spot is covered. The wide input area of the FAST detector, very useful for the large area systems, becomes a problem for this type of experiment as much more detail is needed. To overcome this a slit is used to mask off all but the required area of the NaI detector. The height of the slit allows for up to 3mm of vertical beam spreading to occur in the atmosphere and still be incident on the detector and hence be included in a single reading.

Typically a 0·01° step is used, covering 0·5° to 1° in both 2θ and ω. In order to prevent saturation, the experiment uses 10 seconds per position and this leads to an average of about 10 hours per set of readings.

If the scintillation detector were to be replaced with an area detector the resultant image would be as represented in figure 2.3a. The x-axis of this image is generated by the angular rotation in 2θ. The y-axis is not a feature of the crystal structure but is generated by atmospheric effects and crystal irregularities. This is effectively integrated by the scintillation detector into a single one dimensional image representing intensity in the 2θ axis, as shown in figure 2.3b.

After a single $\Delta\omega$ rotation is done, another such intensity profile is generated. Eventually an intensity picture of ω against 2θ is produced and a corresponding contour map can be calculated. This represents the X-ray densities within the reciprocal space



(a) The detected image.

(b) The resultant line.

Figure 2.3: The integration from an area to a line.

of the reflections.

Many problems are faced when doing this type of experiment as the required information is often swamped by spread in the wavelength, λ, of the radiation and non-uniformaties in the crystal structure. Figure 2.4 shows how various effects contribute to the contour map. The mosaic spread, denoted by μ, is the effect of crystal irregularities and causes a spread in the ω axis of the $\omega, 2\theta$ space. The source variations, σ, are always parallel to the θ axis and are caused by the changes in the radiation beam coming into the crystal. The output beam wavelength, which is the only feature dependent on θ, causes changes in the λ component.

2.2 X-ray Detection Using Charge Coupled Devices

2.2.1 A phosphor based, continuous, X-ray photon detector

There have been a few attempts to use phosphor screens to convert the X-rays into optical photons. Examples of these include X-ray microscopes [24], image analysis [25] and some work in the crystallography fields of protein [26, 27] and synchrotron [28] studies.

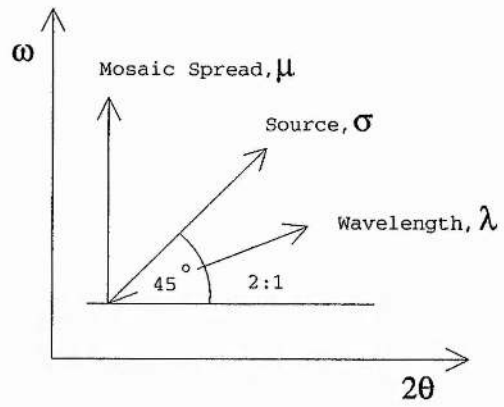


Figure 2.4: The components of the reciprocal space spot profile.

One of the main problems to be overcome in the single Bragg reflection analysis is the need to detect as many of the X-ray photons as possible to reduce the experimental time. The conventional drive electronics for a CCD has a frame shift period that would cause problems as the correct position of incident photons would not be recorded during such a frame shift. Before looking at direct illumination of CCDs by X-rays, an idea developed in conjunction with Dr. D.S.S. Robb at St. Andrews university in 1988 will be discussed. This system uses a phosphor detector feeding into the CCD image collectors.

Figure 2.5 shows the schematic of this proposed detector. Its operation is very simple: the X-ray photons are converted to optical photons at the phosphor, the photon beam is passed through a gain stage and then focussed onto a beam splitter. The two beams now go to two CCDs via shutters that will alternate the beam onto first one CCD and then the other. While the beam is active on one CCD the other is being read out and vice versa. This ensures continuous analysis of the incident X-rays.

There are clearly many problems of aligning such a system and the two output signals would also have to be combined to get the correct real time signals.

Normal phosphors can give close to 100% quantum efficiency (QE), with a 1·5$\AA$ X-ray photon producing 3000 4500$\AA$ optical photons. These are emitted into 4π steradians. In normal conditions however, this would normally lead to only 300 photons coming out in the correct direction (a $0{\cdot}5\pi$ cone). This coupled with the losses in the system and in a CCD, where a QE of only 80% can be expected, means a large gain stage is required to operate this experiment.

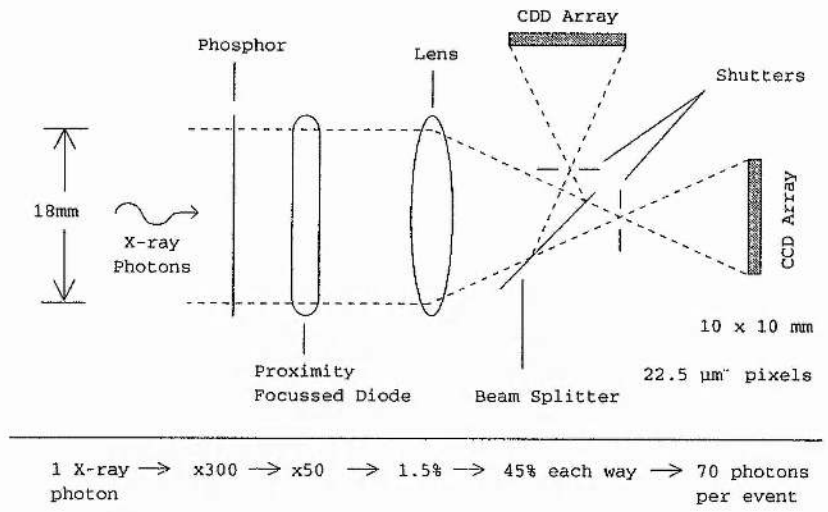


Figure 2.5: A real time continuous photon detector.

One way to improve this is to use some form of containment for the phosphor. Small fibre optic bundles can be hollowed out for the first few millimeters and can then be filled with phosphor [29]. This means that all the light emitted in the phosphor is channeled down the fibre it is emitted from. This reduces the resolution of the detector but does form a good solution to the containment problem. The layout of such a system is shown in figure 2.6. It is also possible to use small micro tubes filled with a scintillating material core [30] instead of fibre optic lines.

In order to get the intensity signals up to a suitable level for the optical response of the CCD, a proximity focussed diode is required. This is simply an optical diode amplifier that has the emission stage very close to the gain stage ensuring that the possible spread of photons is very small across the unit and hence good resolution is maintained.

There is also the problem of having very low incident X-ray counts. This would mean that the image could not simply be treated as an intensity map and some interpretation of the centre of each light cone (created when the light spreads out as it passes through the apparatus) would have to be done so as to identify the position of the individual incident X-rays.

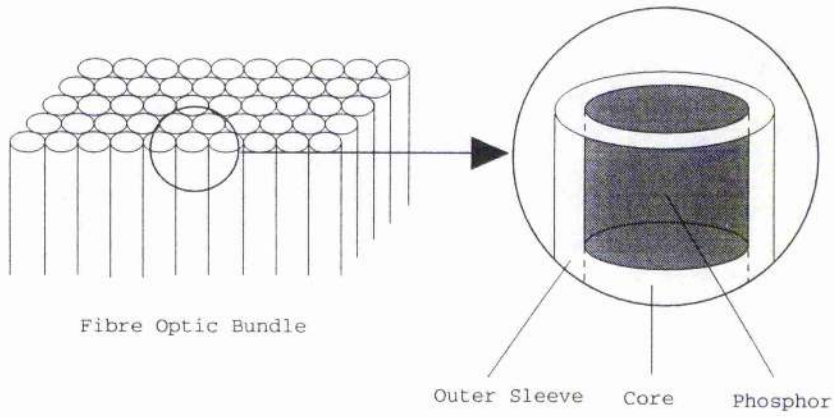


Figure 2.6: Containment of emitted optical photons.

2.2.2 The suitability of CCDs to X-ray applications

Unfortunately this device would not achieve the vertical integration, done by the slit in the original apparatus, that is needed to handle the vertical spread of the Bragg reflections, though it would provide a very good alternative to the existing SIT tubes. Because of this limitation a new system had to be developed that would allow vertical integration and also reduce the complexity of the instrument as a whole.

One development in CCD technology is the ability to directly irradiate them with soft X-rays and allow the silicon to provide the required gain [31, 32, 33]. The data given in table 2.1 provides some useful insights into the response of EEV CCDs to X-ray photons [34, 35]. This is clearly to be preferred as the CCD has an extremely good point spread function and also provides very good rigidity, immunity to magnetic fields, and very well defined structure that eliminates the need for complex distortion calculations and thermal compensation circuitry. The CCD is also a much smaller unit than the existing SIT tubes, and other opto-electronic devices such as the photicon [36] and vidicon [37] cameras and would thus lead to a smaller, more versatile instrument.

From the data in table 2.1 it is possible to show the following:

$$\text{Saturation} = 300\text{nA} \times \frac{1130}{1\text{nA}} = 339 \times 10^3 e^-,$$

$$\text{Energy of Cu } \alpha \text{ line} = \frac{hc}{e\lambda} = 8{\cdot}05 \times 10^3 eV;$$

CCD peak output voltage	180mV
Peak output current	300nA
1nA output per cell	$1130e^-$
λ_{Cu}	$1{\cdot}54\mathring{A}$ (α line)
No. (e^-, e^+) per detected X-ray	$\sim \frac{\text{E (eV)}}{3{\cdot}6}$
QE at $1{\cdot}54\mathring{A}$	32%

Table 2.1: Data on the P86320 CCD.

$$\text{Gain} = \frac{8{\cdot}05 \times 10^3}{3{\cdot}6} = 2240 \text{ @ QE of } 32\%;$$

$$\text{Saturation} = \frac{339 \times 10^3}{2240} = 151 \text{ X-rays};$$

$$1\text{ADU} = \frac{151}{255} = 0{\cdot}59 \text{ X-rays},$$

$$= \frac{339 \times 10^3}{255} = 1330e^-.$$

Clearly it is possible to irradiate an area CCD with the X-ray beam directly. The CCD cells will collect the incident photons and, as the frame is read out it can be digitized and the digital rows summed to achieve the required vertical integration. The CCD provides good QE at soft X-ray wavelengths [38] and studies have shown they are very suitable to general X-ray detection [39, 40, 41] and even to single photon counting [42]. Figure 2.7 shows the quantum efficiency of CCDs at non standard wavelengths.

One other problem that must be considered is the build up of dark current relative to the required detection levels. At its worst the dark current at 21°C reaches 3nA. From the data this corresponds to

$$3\text{nA} = 3 \times 1130 = 3390e^-,$$

which in terms of suitable ADC units corresponds to 2·55 counts. When compared with the actual signal this presents a problem as it will swamp any wanted, X-ray generated, electrons. Chapter 3 will cover the methods developed for measuring and then subtracting this dark current signal. The error that must be considered here is due to the random nature of the dark current. This is dependent on the square root

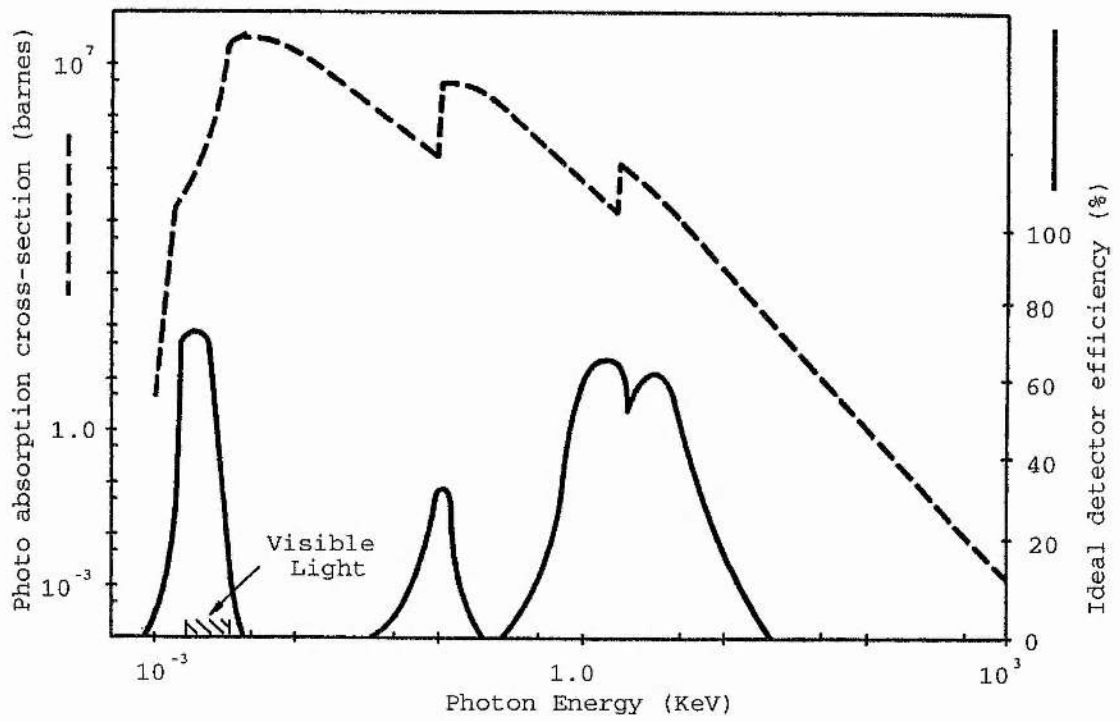


Figure 2.7: Quantum Efficiency of CCDs.

of the number of thermally generated electrons. Hence the maximum error is given by

$$\text{Error} = \sqrt{n} = \sqrt{3390} = 58 \text{ electrons.}$$

Where n is the number of electrons involved.

From this the fractional error in ADUs can be seen as

$$\text{Error} = \frac{58}{1330} = 0{\cdot}04$$

i.e. the total dark current at maximum will be $2{\cdot}55 \pm 0{\cdot}04$ ADUs. As 0·04 is much less than the count generated by the incident , X-ray, photons, the subtraction will easily produce an accurate detection method.

There are of course limitations to the use of directly illuminated CCDs. Radiation damage of the electrodes and the silicon itself has been detected in many devices. Technical publications by EEV [43, 44] suggest that, if back illuminated devices are used, then the high resistivity devices can withstand soft X-ray radiation indefinitely for intensities of less than 10^4 rads.

Consultation with EEV suggested a number of ways of improving the application of CCDs in this area. The use of higher resistivity silicon [45] and depletion layers greater than 30μm [46] should produce much better QE. Though expensive (a $100\Omega cm^{-1}, 100\mu m$ device currently being around £6000), such devices are now becoming more readily available and are provided in the same format as the cheaper optical units. Because of this the following experiments were carried out on normal optical CCDs and when future work requires it, the more suitable X-ray sensitive devices could be substituted. Once again it is in the field of astronomy detectors that much of the pioneering work in high sensitivity CCDs has been carried out and has led to a number of relevant device evaluations [47, 48].

Research work at EEV is being performed in many of these fields. In particular they are developing a low dark current, back illuminated device for £1300, and work using $31k\Omega cm^{-1}$ silicon is producing good results. They are also developing 11μm devices for high definition TV systems. Unfortunately at the time this research was being carried out none of these devices were available but, as they were being designed to be electrically identical to the normal TV sensor, the development work described here should be applicable for later devices.

In order to apply the area technology of the CCD to Prof. Mathieson's experiment, it is necessary to find some way of vertically integrating the pixels to simulate the use of a slit and a scintillation counter. One possible solution would be the use of a large pixel linear array of sensitive elements. Such arrays can have pixels with the same width as the area CCD pixels (22μm) but with a height of up to 5mm. These linear detectors are often known as reticons[†] and are based on either CCD or diode array technology [49].

These were initially developed as position detectors using phosphor [50] or phosphor and fibre-optic [51] couplings. Development continued into the directly radiated sensors for crystallographic applications [52, 53, 54], spectroscopy [39, 40, 55] and then into synchrotron detection [56]. The main problem that these units suffer from is that the noise levels, particularly the dark current, are dependent on the area of the cells and hence when large cells are used the noise levels go up. As diode arrays are inherently more noisy that CCDs anyway, they become unsuitable for the lower intensity

[†]The company EG&G Reticon pioneered work into these detectors.

applications. They also have the same dynamic range as units with smaller cells. This means that the beam intensity (per m^2) must be lower as the larger area pixels will reach saturation faster than would devices with smaller pixels.

This leads on to the application of area CCDs to one dimensional detection. As the dynamic range of the CCD cells is the same as that of similar linear CCD detectors it is possible to use the cells of the area detector to trap the charge, and then 'add up' the charge after reading out the individual values. In this way a dynamic range of 288 times the linear celled devices can be obtained for a normal TV sensor of 388×288 cells.

If an area CCD was to be used this vertical summing could be done in two ways. If the CCD were to be mounted on its side, then as a 'row' was read out, all the pixels could be summed one after the other and then at the end of the 'row' a resultant value could be collected. This would be an ideal solution except for the fact that it is necessary to have continuous data collection. When a frame is shifted out of the active area of the CCD and into the storage area it will pass horizontally through the beam if the CCD is on its side. This means that the more complex problem of having to vertically sum pixels when the CCD is mounted normally has to be solved. When the CCDs active area is shifted down, the incident photons will still appear in the correct horizontal position and, as the hardware is vertically integrating, the photons vertical position is not important.

Because of the low intensity of incident X-rays on the CCD it is very improbable that there will be more than one X-ray photon per pixel per frame (assuming a normal 50Hz frame rate). Because of this it is only necessary to use a logical comparison on the output to detect events. Such a circuit will look at the charge on the output of the CCD and compare it to the charge that would be expected if a single X-ray photon was incident on that cell.

Within Prof. Mathieson's experiment a 0·1mm slit is used giving a peak count of about 10000 events per second, spread over the full 5mm height of the detector. As the pixels are 22·5μm high, this ensures that over the area of the original slit a CCD detector would see

$$\frac{\text{No. Counts}}{\text{No. pixels}} = \frac{10000}{\frac{0{\cdot}1\times10^{-3}}{22{\cdot}5\times10^{-6}}\frac{5\times10^{-3}}{22{\cdot}5\times10^{-6}}} = 62{\cdot}5 \text{ X-rays per pixel.}$$

In normal operation a TV sensor is running at 50 frames per second and hence an

expected $\frac{62.5}{50}$=1·25 X-rays per pixel per frame should be seen. This means that, if the apparatus were to be run at 80% maximum power, the worst case would be to expect one X-ray photon per pixel.

Clearly such a setup will lose some of the events. Either a pixel will contain more than one incident photon, in which case the extra charge will have to be recognized and suitably processed, or the photon will appear at the edge of a cell and the charge will be shared between two cells. It has been shown that, in the worst case, 50% of incident photons cause some energy to be spread between adjacent pixels [40]. This can be removed by computer processing of adjacent pixel intensities but it certainly identifies the need for some type of post processing of the results. There have been a number of attempts to do this type of processing and algorithms are generally available. For this application a standard method using a 3×3 grid [48] to process the energy readings would be ideal but, in order to analyse simple data a direct voltage comparison will be used here.

The hardware will have to store a complete row of data at any one time. Each pixel voltage read out of the CCD will have to feed into a comparator and be converted to either a '1' or a '0' depending on whether it is above or below the required charge level for a single detected event. This will then be added to a stored value for that horizontal position and the resultant sum held until the next pixel is available for this horizontal position. At the end of the frame, and within the vertical flyback time, all the 388 values representing the vertical sum must be read out into the analysing computer system.

2.3 Vertical Integration Hardware

2.3.1 Basic design principals

If the X-ray camera were to be run at the normal TV rate of 50 frames per second the pixel information would be available at a rate of 8MHz. This means there are only 125ns available to do the read out, increment and write back, of the vertical summed data. The timing of this is very critical. The currently available memory chips have 35ns access times. This, added to a counter update time of 25ns, the incremental time of a PAL summing device, 20ns, and latch delays of 10ns or so means that there is no

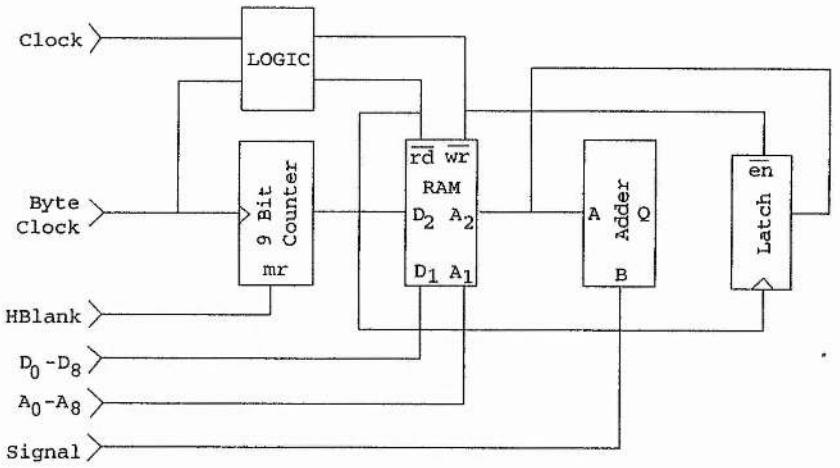


Figure 2.8: Block diagram of the vertical summer hardware.

spare time available

The most obvious consequence of this is that the memory device has to be a dual port unit. The added delays of multiplexing the data and memory busses with the processor would prevent the timing constraints from being achieved.

As can be seen from figure 2.8 the remainder of the hardware is relatively simple. A counter is updated from the pixel clock and provides the addresses for the dual port memory device. This in turn reads out a value into a summing section where the comparator output is added to it. The logic then latches the resultant output and writes it back into the memory. This process continues for the whole line, after which the counter is reset and the process begins again. When the vertical blank signal is detected by the processor a DMA cycle is initiated that reads out the data and then replaces it with zeroes ready for the summing of a new frame.

2.3.2 Summing via programmable logic

Because of the very short time available to the summing process, traditional synchronous adders are not fast enough. Alternatives were considered. Firstly a loadable counter could be used, feeding in the pixel value and then adding one to it if necessary, or secondly, a logic array could be used to add in the value asynchronously. The timing complexity of the loadable latches lead to unnecessary complications as did the need to use multiple devices (loadable counters such as the 74LS191 are only available in 4

bit or decade counting packages and then not always in the faster logic types such as the 74F family).

The 16v8 GAL gives the flexibility needed and is also available in a 15ns access version. This, though expensive, gave better timing than the TTL versions could, and also provided the flexibility to do other logic functions within the circuit.

The basic equation of carry forward adder is found by extending the half adder logic,

$$D_0' = D_0 \oplus B, \tag{2.1}$$

$$C_1 = D_0 \wedge B, \tag{2.2}$$

where D_0 is the least significant bit, B is the input bit and C_1 is the carry produced by this sum.

This can be extended so that

$$\begin{aligned} D_0' &= D_0 \oplus B, \\ D_1' &= (D_0 \wedge C_1) \oplus D_1, \\ &\vdots \end{aligned}$$

This set of equations can be substituted with equations 2.1 and 2.2 to get

$$\begin{aligned} D_0' = \quad & D_0 \wedge \overline{B} \\ \vee \quad & \overline{D_0} \wedge B, \end{aligned}$$

$$\begin{aligned} D_1' = \quad & D_0 \wedge B \wedge \overline{D_1} \\ \vee \quad & \overline{D_0 \wedge B} \wedge D_1, \end{aligned}$$

$$\vdots$$

In order to get this into a form suitable for a logic array, these equations must now be expanded to get

$$\begin{aligned} D_0' = \quad & \overline{B} \wedge D_0 \\ \vee \quad & B \wedge \overline{D_0}, \end{aligned}$$

$$\begin{aligned} D_1' = \quad & B \wedge D_0 \wedge \overline{D_1} \\ \vee \quad & \overline{D_0} \wedge D_1 \\ \vee \quad & \overline{B} \wedge D_1, \end{aligned}$$

$$\vdots$$

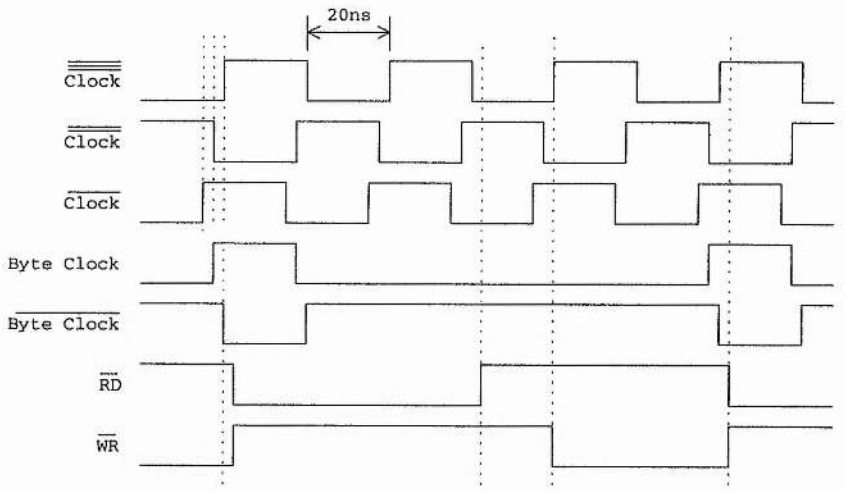


Figure 2.9: Timing of the read and write pulses.

Because of the need for 9 bit output, and also due to the limited number of minterms available to the array, it was necessary to use a further array to generate the most significant bit. Thus both D_7 and D_8 were defined in a second PAL. The data was fed into these synchronously by providing a carry output from the first PAL, defined as

$$C = B \wedge D_0 \wedge D_1 \wedge D_2 \wedge D_3 \wedge D_4 \wedge D_5 \wedge D_6,$$

i.e. the carry output of bit six. The definitions of D_7 and D_8 were then just the same as for D_0 and D_1, but with this carry bit being fed in instead of the comparator output.

2.3.3 Generating read and write line timing

The timing generation for the memory had to be derived from the CCD clock and the byte clock (as defined in chapter 1). The CCD clock provides a higher frequency clock that can be reduced to provide the correct read, write and control pulses to drive the circuit.

Figure 2.9 shows the required timing for this circuit and includes the slight discrepancies in the pulse generations that the gate delays produce. First the data is read out, with $\overline{\text{RD}}$ active, (low). Then there is a delay to allow the addition to be done and finally the data is latched and fed into the memory with the $\overline{\text{WR}}$ signal low. Normal design techniques would lead to either a state machine design, or to a cascaded flip-flop system. Unfortunately these would incur too many time overheads and would not

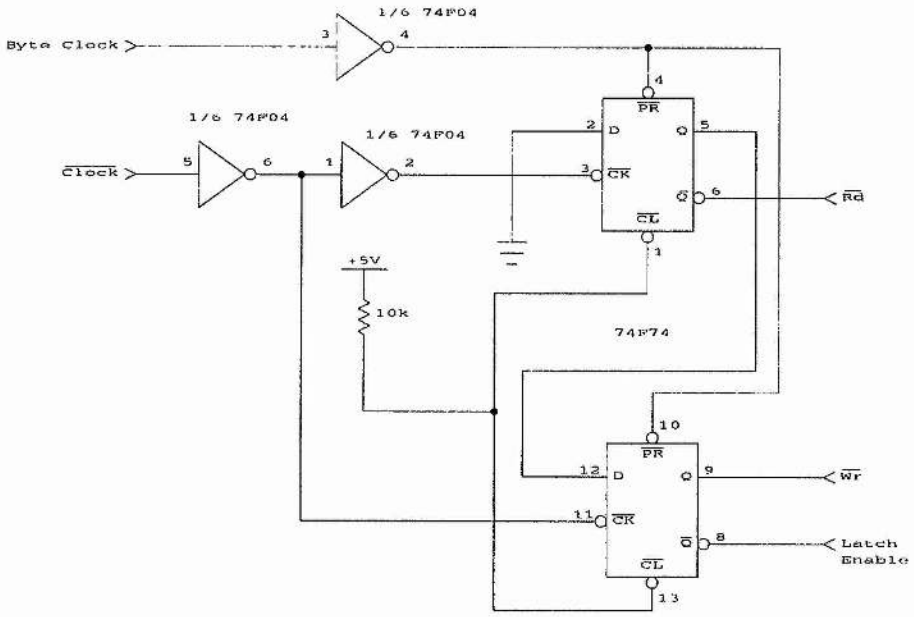


Figure 2.10: Circuit element showing the read, write and control timing.

provide the exact timings needed for this application. One solution is to make use of the finite rise times of a flip-flop output.

Figure 2.10 shows the circuit element used to implement this timing. It works by letting the pixel clock reset the outputs to their initial positions of $\overline{\text{RD}}$ low and $\overline{\text{WR}}$ high. The next falling edge of the CCD clock comes just before the end of the byte clock since the byte clock is derived from the CCD clock. This means that this edge is ignored by the flip-flops. The second falling edge causes the $\overline{\text{RD}}$ signal to be set to high as a 0 is clocked into the flip-flop from the data input.

This high value is now clocked into the lower flip-flop by the next rising edge of the CCD clock and hence the $\overline{\text{WR}}$ signal is activated. Finally the whole circuit is reset by the pixel signal when it goes active again at the start of the next pixel output cycle.

In order to make use of the propagation delay time in the middle of this timing period, the $\overline{\text{Q}}$ output of the lower flip-flop was used to latch the data signals driving the inputs of the PALs.

Though this circuit operated well, it was, not surprisingly, very unstable as a strip-board circuit and any loading of the flip-flops caused timing changes. This meant that

Device	Price†	Access Time	Size
IDT 7130	£13·58	35ns	1k× 8
IDT 7025	N.A.	35ns	8k× 16
PDSP16520‡	N.A.	50ns	1k× 16
VT16DP8	N.A.	70ns	2k× 8
HM63021§	N.A.	28ns	2k× 8
VT16DP8	N.A.	70ns	2k× 8
MT42C8127¶	N.A.	100ns	128k× 8
IDT 70104	£18·62	25ns	1k× 9
	£14·90	35ns	1k× 9
IDT 7010	£17·87	25ns	1k× 9
	£17·87	35ns	1k× 9
MK41H78‖	N.A.	20-25ns	4k× 4

Table 2.2: Dual port memory devices.

a low impedance oscilloscope could not be used as a test device. Once a printed circuit board had been manufactured the stability was greatly improved and the timing became very reliable. Experimentation showed that a 74LS74 was not suitable, creating poor edges on the drive lines, the 74F74 worked well though and provided reliable results, as shown in section 2.4.1.

2.3.4 The choice of memory device

When considering the choice of memory device certain parameters could be defined. Firstly it had to be a 9 bit device because, in the worst case, there may be a photon in every vertical position, leading to a count of 288 photons, requiring 9 bits to store.

Secondly, the device had to have at most 45ns access time as the timing diagram of the read and write lines illustrates. Table 2.2 show a list of the parts considered and their characteristics.

Some of these devices were arranged so as to be stackable, for example the IDT70104 is the slave version of the IDT7010. These devices can be used in parallel, providing the

†As of June 1991.

‡This device is a quad port memory.

§A line memory device.

¶A serial and parallel accessed device.

‖This device also has a built in auto clearing function enabling all the memory locations to be zeroed via a single control line.

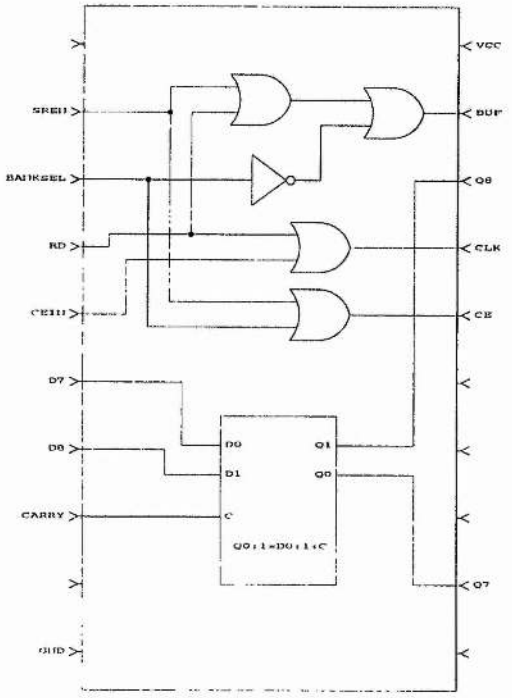


Figure 2.11: Internal layout of PAL 2 logic chip.

larger bit sizes, and allowing the master unit to arbitrate as to which part is accessing the storage. In the hardware developed here, the IDT7010, 35ns part was used [57]. This was a faster device than was necessary thus guaranteeing the timing would be met, and also proved to be the most cost effective of the devices available. Being a 9 bit device (designed for parity checking of 8 bit words) it provided everything that was needed. In the event that a faster version might have been needed it could also have been upgraded to the 25ns version.

2.3.5 The final circuit

Inserted figure IV shows the circuit diagram of the final hardware implementation. The logic arrays PAL 1 and PAL 2 provide the summing described earlier, and the remaining part of PAL 2 was used to provide most of the less time critical logic functions (figure 2.11). Two positive edge triggered latches, a 74F374 and a 74F74 were used to latch the 9 bit output of the summer, a 74LS125 buffer was also needed to add the $\overline{OE}$ feature to the flip-flop.

There was also a need to modify the original timing requirements because the update rate of the 74HCT4040 was still not good enough to meet the timing. Two latches (74LS374s) were put in to hold the counter output one cycle behind the update clock. This removed the dependence on the update time of the counter but no longer allowed the correct memory accessing timing, having the effect of delaying the entire image by one pixel and dropping one pixel value from the end of the line. The 388 pixels per line are effectively reduced to 387. As this instrument was developed to provide an experimental confirmation of the principles, this was thought to be a reasonable compromise.

2.3.6 Accessing the RAM

The other problem with the dual port RAM was in getting the 9 bit data into the processor storage RAM. The read cycle of the dual port RAM produced all 9 bits of stored data on the 9 data lines. By latching the most significant bit into a flip-flop and enabling it onto the 8 bit processor data bus via a tristate buffer, it was possible to ensure that after reading the RAM a subsequent read could be from this latched bit and thus all 9 bits of data could be collected.

Unfortunately the use of a DMA chip presented problems as the DMA is not capable of this kind of complex switching. The solution was to use the lowest address line, A_0, instead of a selection line and then address the RAM chip with address lines A_1 to A_9. Now, even bytes contained the main 8 bits of the word and the odd bytes contained the one bit data that corresponded to the ninth bit of the previously read word.

It was now possible to use the DMA to zero the RAM contents provided this ninth bit was suitably dealt with. When a zero byte was written to the RAM the ninth bit was pulled to ground ensuring that all nine bits were written as zeroes. This doubled the time the DMA took to erase the bytes because the addressing was now such that every odd byte acted the same as every even byte (for write cycles). As will be seen from the readout timing, this increase was just acceptable.

2.3.7 Achieving high speed data transfer

In order that the DMA transfer can be done as fast as possible, two areas of memory within the Z80 address space were set up. One contained zeroes for 776 bytes. This

was used to allow the DMA to erase the dual port memory by simply transferring this blank data into the dual port address space. The other section was used to store the data obtained from the DMA transfer when the dual port memory was read out.

Given a clock frequency for the DMA of 4MHz, a transfer rate of one byte per 6 clock cycles, the need to transfer 776 bytes into the Z80 RAM and then the need to erase 776 bytes to zero, the timings of the readout system are found to be

$$\text{transfer cycle time} = \frac{6 \times 2 \times 776}{4 \times 10^6} = 2{\cdot}33\text{ms}.$$

If the register setup times for the DMA are included with this, an expected cycle time of 2·4ms would be reasonable. This can be compared with the same timings using the LDDR Z80 command which requires

$$\frac{21 \times 2 \times 776}{6 \times 10^6} = 5{\cdot}43\text{ms}$$

which, as the flyback time was measured to be $3{\cdot}07 \pm 0{\cdot}02$ms, could not be used. Clearly the DMA solution is just capable of achieving real time data movement.

Having ascertained that the proposed method would work, the software was then implemented via the following algorithm.

```
1. Fill Z80 RAM area with zeroes.
2. Setup DMA transfer parameters.
3. Wait till Vblank signal goes active.
4. Do DMA Transfer of dual port data memory into Z80 storage memory.
5. Setup new DMA transfer parameters.
6. Transfer zeroed memory area from Z80 into dual port RAM.
7. Transfer stored data from Z80 to PC.
   (Most of this happens during the frame period).
8. Repeat from 2 until required number of frames have been summed.
```

Once this had been implemented, software on the PC was generated to take in each set of values and store them. Options to display, graphically and numerically, the data from any particular frame were also provided. With a 4M byte RAM PC, about ten thousand frames can be stored, easily enough to ensure that no hard disk transfers would be required. This is necessary as these would slow the data collection process down were they to be needed during the experiment. Data can still be written to the hard disk after the experiment has concluded.

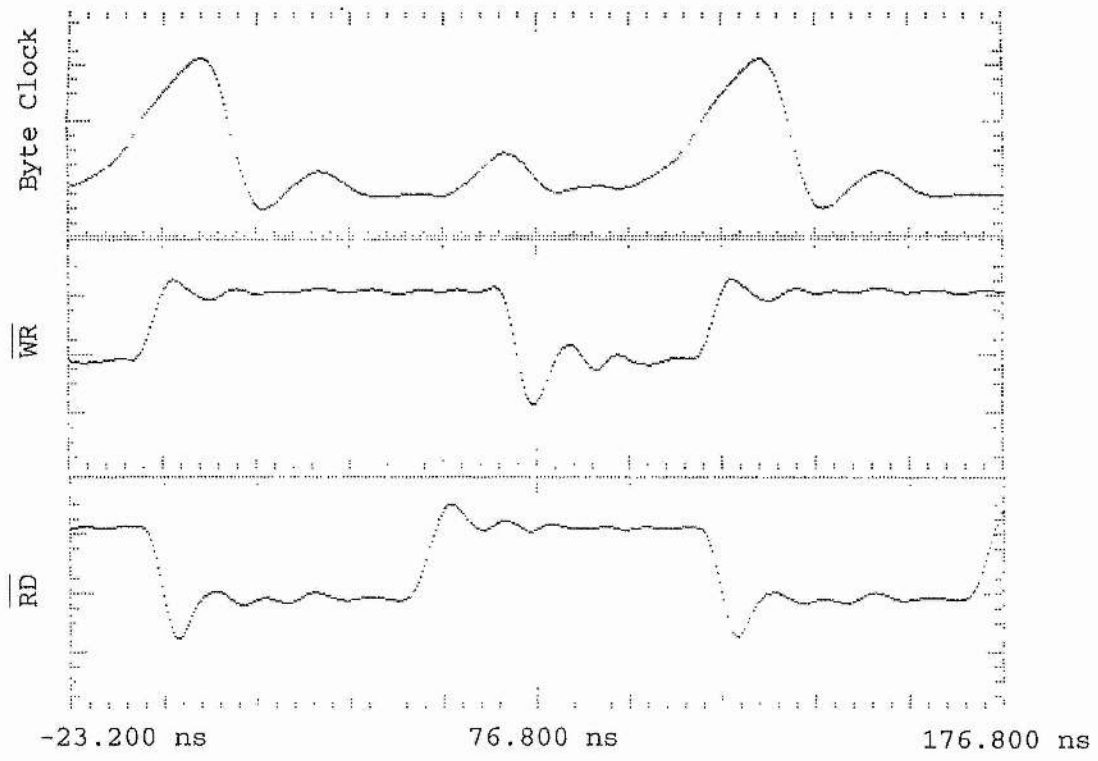


Figure 2.12: Measured output of the read and write enable circuit.

2.4 Results

2.4.1 The read and write signal timing

In order to verify the output of the memory control hardware, it was necessary to use a fast, high input impedance oscilloscope. This was mainly due to the fact that the circuit relied on the fast edges of the flip-flop outputs and, when a capacitive load was applied to these outputs, the edges became less sharp and the timing shifted enough to become unstable.

Figure 2.12 shows the clock inputs along with the $\overline{\text{RD}}$ and $\overline{\text{WR}}$ outputs from this circuit as measured by a Hewlett Packard 1GHz digital sampling scope.

Comparison of these timings with those given in figure 2.9 illustrates that this circuit produces the required pulse durations very accurately.

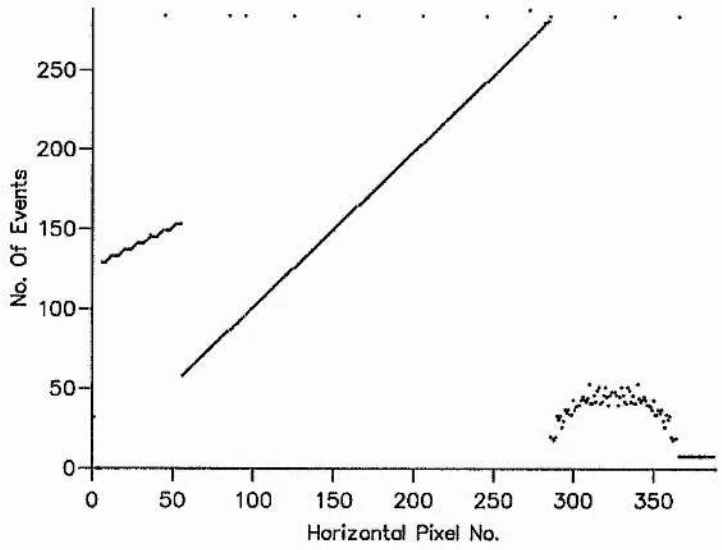


Figure 2.13: Hardware summer acting on the test image.

2.4.2 Vertical summing in software

In order to test the hardware, it was necessary to have a comparison. By accessing the graphics ability of the hardware developed in chapter 1, and not attempting this in real time, it was possible to use either a test image or to read a frame grabbed image into the PC and write a program to compute the correct output of the vertical summer. This program took 123 ± 1s to execute when run on a simple test image.

2.4.3 Hardware results

Figure 2.13 shows the output of the summing hardware when operating on a test image consisting of a filled circle (right hand side), a filled triangle (middle) and a shaded vertical bar (left hand side) all superimposed on a grid using 40 pixel spacing. This compared very well with the output of the software summer.

Once positive results had been obtained, the hardware was run on a frame grabbed image (figure 1.13), using the most significant bit of the output as the summing circuits input data bit. This result is shown in figure 2.14. Finally the system was run in real time using the live video signal. The results, shown in figure 2.15, agree very closely to the expected values and even show some extra noise immunity over the frame grabbed

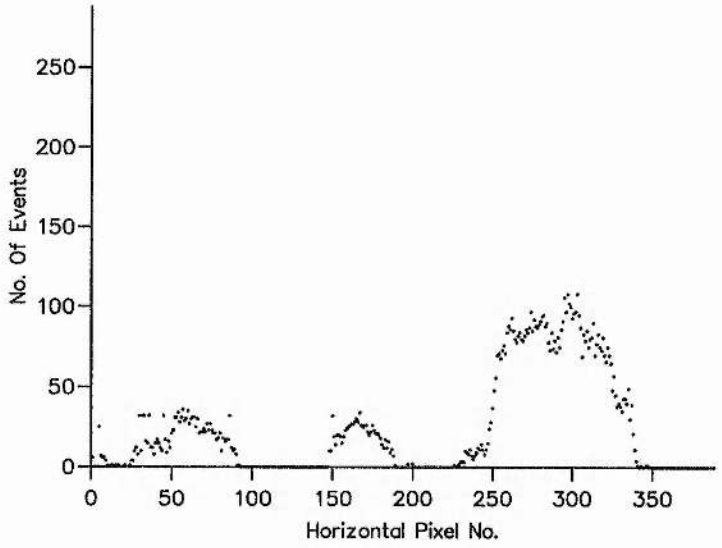


Figure 2.14: Summing a frame grabbed real life image.

version as the camera system was not loading the video signal.

2.4.4 Future work and radiation testing

Before work can be continued in this field, it will be necessary to build the entire system onto a printed circuit board, and preferably go to surface mount devices to reduce the noise levels. Redesigned power supply noise suppression and very good decoupling would also be beneficial. This should allow the noise floor to be lowered below the requirements on the X-ray detection circuitry. It would also be necessary to provide active feedback to the comparator to allow the dark current to be subtracted.

Testing of the camera on X-ray sources proved very difficult because of the camera noise levels. The experiment uses radiation from either $1{\cdot}5\mathring{A}$ copper (Cu) or $0{\cdot}7\mathring{A}$ molybdenum (Mb) sources. These have an energy given by

$$\begin{aligned}
\text{energy (eV)} &= \frac{h\nu}{e} = \frac{hc}{e\lambda}, \\
&= \frac{3 \times 10^{8} \times 6{\cdot}626 \times 10^{-34}}{\lambda \times 1{\cdot}6022 \times 10^{-19}} = \frac{1{\cdot}241 \times 10^{-6}}{\lambda}, \\
&= \begin{cases} 8{\cdot}271\text{keV} & \text{Cu @ } 1{\cdot}5\mathring{A} \\ 17{\cdot}72\text{keV} & \text{Mb @ } 0{\cdot}7\mathring{A}. \end{cases}
\end{aligned}$$

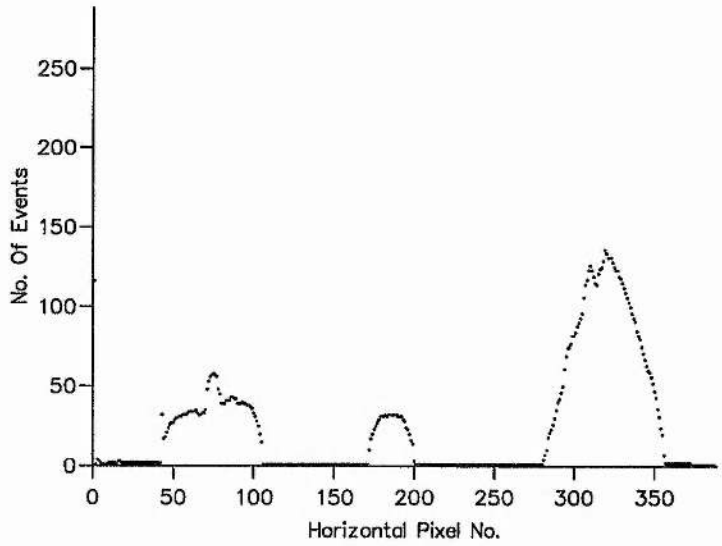


Figure 2.15: Summing a live image.

Isotope	Energy
Cobalt (Co_{60})	1·173, 1·333 MeV
Caesium (Cs_{137})	22 - 26 keV
	88 keV
Manganese (Mn_{54})	5·4 - 6·0 keV
	835 keV

Table 2.3: Available radiation sources.

Table 2.3 shows some possible γ emitting test sources [58]. The Mn_{54} should be of particular use as it has radiation levels well into the normal CCD sensitive regions, giving a 10-15% quantum efficiency (rising to 60% for a 60μm silicon CCD). The source that most closely matches the X-ray source is an Fe_{55} isotope. This was not available for use however so tests could not be done at this time.

Chapter 3

A Solid State Streak Camera

3.1 Introduction

3.1.1 The generation & use of short duration laser pulses

For some years now it has been evident that short duration laser pulses have many applications. With mode locked dye lasers [59] now capable of producing reliable pulses with durations in the low femto-second ranges [60], it is now possible to create 'packets' of light only a fraction of a millimeter long. Within such time scales molecular interactions can be filmed as sequences of images [61], tectonic plate movements can be detected to an accuracy of a single millimeter via satellite ranging [62] and optical properties of solids can be very accurately determined using the broadening effects on the monochromaticity of such pulses.

Along with an increase in the numbers of diode lasers and mode locked systems, there has also been a build up in the number of medium length pulse sources. These have medical applications as they are able to deliver large amounts of energy in a very localized way but cause minimal heating as the pulses are short compared to their 'off' periods. There are also telecommunications uses for medium length pulses as they can be used to modulate digital data onto a fibre link.

One problem that all short pulse research suffers from is that detailed analysis of the pulses is hard to do. Power meters such as the Newport 835 or 840, can look at changes with laser output power with time but cannot look at mode changes or frequency broadening within the pulses.

If more advanced research into these pulses is to be achieved, one instrument that is needed is a fast detector that can look at the beam profile in time, and not the

integrated power. For medium pulse duration lasers this would need to look at a period of, say, 200μs in 10μs frames. With an emitted beam profile of about 5mm such a camera would need to be able to resolve 50μm detail and be able to cover the entire beam width in one frame.

3.1.2 Existing measuring techniques

There are a number of techniques currently in use to analyse laser pulses. Sensors such as light dependent resistors are too slow for these applications but some of the faster optically sensitive silicon detectors can be used, allowing longer duration pulses to be looked at with light sensitive diodes similar to those used in conventional power meters.

There are two ways of looking at the faster pulses. Either the beam is passed through a streak camera [63] and looked at in a multiple or one shot recording mode, or an autocorrelated sampling technique [64] is used.

Within a streak camera the laser pulse is converted into an electron beam that is then deflected, at radio frequencies, across a phosphor screen. A very high deflection rate can then be used to trace the beam across the screen and build up an image that can be read out via a photodiode array. For repetitive sampling very good results are obtained down into the pico-second pulse durations. The main problem with such instruments is that they are hard to tune and give poor results when used in a single shot mode. The power gains required to move the beam are achieved by using sharply resonant circuits and these are only tunable over a very short range of frequencies. As the gain is also hard to adjust, very fast pulses cannot be analysed as the low intensity output is not detected by standard readout systems.

The alternative for very fast pulses, in the femto-second region, is to use autocorrelated sampling technique where a composite image is obtained by instantaneously sampling the magnitude of successive pulses each at a slightly later time than the last. In this way all the parts of the pulse are examined and, assuming that all pulses are identical, the shape of the pulses can be determined.

One problem with all these detectors is that they examine the whole beam profile rather than only one single part of it. This means that they integrate the power of the beam over its whole cross sectional area (see section 3.6.3). There are very few

techniques for looking at beam shapes, or profiles, while the pulse is occurring.

One possible method of doing this would be to use a framing camera [65]. These use very fast beam deflection systems to move an image across the surface of a CRT and hence onto a photographic plate. After each frame period, the entire image is moved to a new location on the plate and a small exposure time is allowed before moving on to the next frame position. This means that after the end of the imaging period a set of frames has been taken. By developing the images and then scanning with a densitometer, the beam profiles can be built up.

This system, while easily fast enough to do 10μs per frame (and even down to 1ns per frame) is not capable of looking at very many separate frames. The simple system usually accommodates only a two by eight grid and hence only 16 pictures can be taken. It is also very difficult to use as the deflection tubes are large and require detailed setting up for any one imaging session. The delay in developing the images is also a problem as it can take many attempts to get the precise setup that will produce useful images. Even so framing cameras are readily available from such suppliers as Hadland, T.R.W and Hamamatsu.

Clearly such a system is not suited to the sort of measurements that are required in order to fully analyse medium and short duration laser pulses. A portable, low interference (the radio frequency scanning of streak cameras can cause significant problems) and fast turn around device is required.

3.1.3 A proposed solid state camera

In conjunction with Dr. D.S.S. Robb an idea was developed to make use of a CCD sensor to achieve fast, one dimensional analysis of laser pulses. This was based around a normal TV CCD sensor and made use of the silicon charge storage as both the optical sensor and as a means of storing data while beam analysis was in progress.

The original research into charge coupled devices at Bell Systems laboratories [66] was as a possible means of making digital serial memories and shift registers [67]. A charge insertion device was used to put charge onto one end of a row of charge coupled cells. The charge in all cells could then be transferred one cell along and a charge sensor at the other end could read out the binary, or even analog, signal from the end cell. Unfortunately the problems of dark current, coupled with the rapid growth of

static and dynamic RAM chips, lead to better solutions to the memory problem but charge sensitive detection went on to be applied to linear array and then area array, light sensitive detectors.

The method of one dimensional analysis used here is based on allowing only one horizontal row of an area CCD to be illuminated by the beam at any one time.

Before the pulse starts, the CCD is continually clocked out to ensure that there is no build up of dark current or background light. A trigger pulse from the laser source stops this clocking thus allowing charge from the pulse to build up on the one exposed row. After a fixed integration time, the charge in the whole CCD is shifted down one line and the next integration period started. In this way a succession of lines are stored in the sensor below the illuminated line.

Once the pulse has died away, the CCD is read out. After a few lines of blank data, the first line of the experiment will have been shifted down into the output register and data collection can begin. Lines are read out sequentially until the entire pulse has been recorded.

Because of the very high sensitivity of the CCD, in some cases it is possible to get single photon detection [68], such an experiment is able to handle this single shot detection method with no problem.

The basic digital readout circuitry for such a camera is reasonably well known but there are two significant points of interest in the system used here. Firstly the coupling of the light to only one row of pixels (22·5μm high) had to be achieved with virtually no light leakage into surrounding pixels, and secondly a correction for A/D distortion and dark current had to be developed. The latter was of particular interest as it was hoped to read out the CCD in a slow scan mode so as to reduce noise and improve the output amplification stages. This would mean that a significant amount of dark current would build up during the period of data storage and would have to be subtracted out.

The proposed system has a number of further applications. As the instrument can be mounted at any angle it is possible to look at many slices through the laser beam and hence three dimensional analysis (x, y,*intensity*) can be done. If pulses are found to be reproducible then the instrument could also be moved up and down so as to look at the full pulse characteristics in both axes.

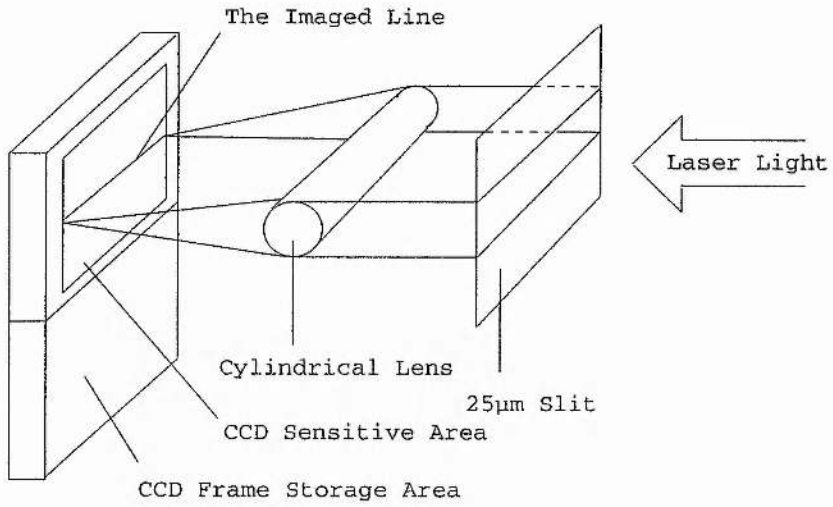


Figure 3.1: Imaging a slit onto the CCD surface.

3.2 Laser Coupling To The Camera

3.2.1 Imaging a slit onto the CCD surface

The first problem to be overcome was the need to couple the laser beam onto a single row of CCD cells. The first attempt was to pass the laser beam through an adjustable slit and then image the first maxima of the diffraction pattern onto the CCD via a cylindrical lens. It was hoped that the second maxima would now fall outside the CCD light gathering area. Figure 3.1 shows the arrangement for this experiment.

A row readout circuit was set up to enable the CCD to be read out onto an oscilloscope and hence allow the output waveform to be analysed. The image was placed vertically on the CCD surface so that the thickness of the imaged beam could be seen as a spread in the row output waveform. If there was any vertical misalignment of the imaged slit then this would be seen as a further spread in the output waveform. To correct for this the slit was adjusted in angle so as to minimize the observed image spread and thus ensure that it was vertical with respect to the CCD pixels.

It soon became evident that the smallest slit that could be clearly imaged did not allow the beam to cover just one row and the alignment was so sensitive that even thermal effects could cause misalignment. The smaller slits were not suitable because the errors in the lens and the non uniform thickness of the slit became too pronounced

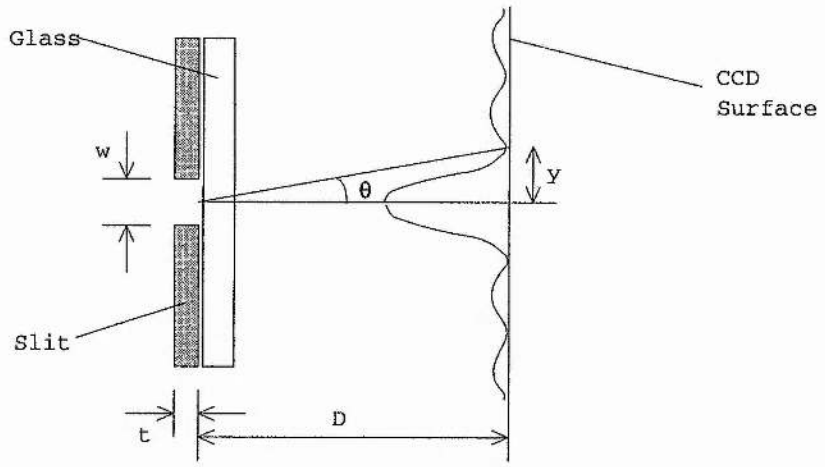


Figure 3.2: The diffraction generated by mounting a slit onto the CCD.

to give useful results.

3.2.2 Mounting a slit in front of the CCD

In order to overcome the problems of the image spread, a razor blade was used, coming down from the top of the image, to form a slit between the blade and the top of the storage section of the CCD. This allowed the laser to directly radiate the pixels without the need for a lensing system.

Testing of this arrangement was carried out by using an existing video camera and simply viewing the TV picture at very high contrast. This revealed that the beam spread, caused by the diffraction pattern being so close to the optically sensitive surface, meant that though it was possible to get the correct beam width at the razor blade, it had spread out by the time the image reached the silicon of the CCD.

The same system was then extended by mounting a Melles Griot 25μm$\times$5mm slit onto the glass front of the CCD. This produced better results as the spread was much smaller with the reduced distance. Figure 3.2 illustrates the effects of diffraction on this arrangement. The distance between the maxima can be found from

$$w \sin(\theta) \approx w \frac{y}{(t+D)} = n\lambda$$

and hence, at n=1, the central peak has a half width given by

$$y = \frac{\lambda(t+D)}{w},$$

$$= \frac{633 \times 10^{-9}(0{\cdot}55 + 1{\cdot}2) \times 10^{-3}}{25 \times 10^{-6}},$$

$$= 44{\cdot}3\mu\text{m}.$$

Clearly the full width of the peak, which contains 90% of the power, spreads over 4 pixels (each with a width of $22{\cdot}5\mu$m).

It was also found that internal reflections within the glass covering contributed to the spread of the beam. The net effect of these problems was to significantly reduce the time resolution of the camera and hence further improvement was needed if the time resolution was to be high enough to meet the requirements.

3.2.3 Coupling via optical fibres

One of the alternative coupling systems available for EEV products is to use a fibre optic block. In this case the CCD-02-06-1-621 sensor was used. This is a scientific grade, TV frame sensor costing about £1170. It contains 6μm diameter fibres coherently bundled to form a 8mm × 5mm block that is mounted close to the CCD surface and very accurately aligns the image on the fibre bundles with the CCD elements. When the 25μm slit was mounted directly onto the faceplate of this fibre block the distance between the slit and the fibre surface was so small as to be negligible and thus removed the diffraction problems. The final version of the optical coupling arrangement is shown in figure 3.3.

The availability of a 6μm slit gave the possibility of going to an even more accurate system were it to be required. This size would also match the 6μm fibre optic thickness.

Alignment of the slit was done on the camera system developed in chapter 1. The slit was placed by hand with a false contrast colour mapping set up to enhance the visibility of any light leakage around the slit. The slit could now be aligned to cover just one row of the CCD and, once a suitable position had been decided upon and the slit fixed using optical wax, the camera hardware could be used to carry out a frame grab of the slit image.

This image could have been analysed to find the vertical position of the slit in a number of ways. The simplest was to measure the height of the image on the screen and the distance from the bottom of the image to the illuminated line. This was done and, given that there are 288 lines on the screen it could be deduced that the slit was,

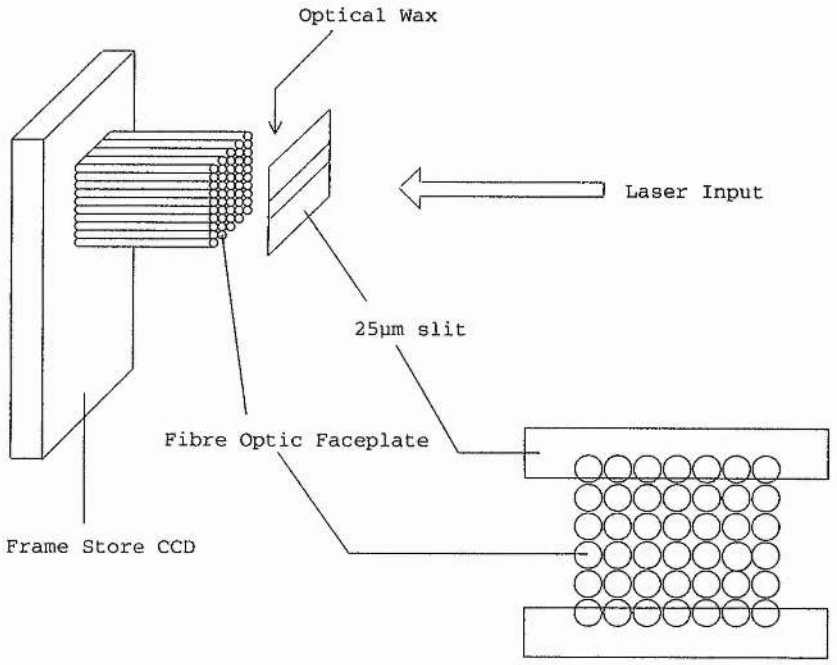


Figure 3.3: The final version of the laser to CCD coupling hardware.

in this case, mounted at line 179. This was added to the number of lines in the frame storage area, 288, to get 467 possible line images before lines of data will be lost in the shifting down process. The hardware described in section 3.3.1 was thus set to a little less than this, 460 or so, to prevent corruption of the initial few lines before the data collection stage begins.

3.3 Digital Controller Electronics

3.3.1 General purpose 'windowing' hardware

The controller for this camera was based around a single digital circuit that was controlled via an 8255 digital I/O board [69] in a PC. The block diagram of this controller is shown in figure 3.4. The clocks were generated from a counter section and then fed out to the rest of the circuit via some enabling logic. Before an experiment starts the fast row clocks are enabled to the row readout drivers and the CCD. These row clocks are counted and, when a predetermined number have been sent to the CCD, the counter is reset and a single pulse is sent to the column shift circuitry. This performs a full three phase clocking cycle on both the frame and store sections of the CCD, shifting

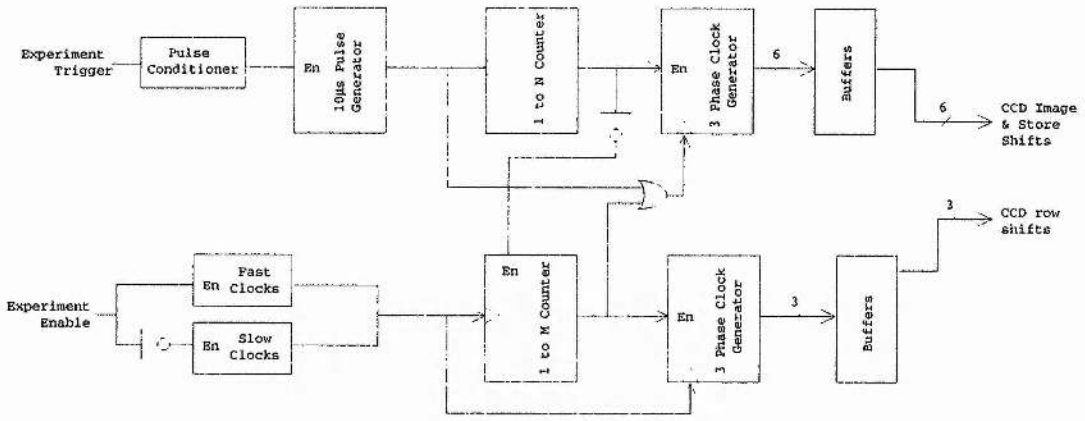


Figure 3.4: Block diagram of the one dimensional camera.

the stored charge down by one row. This process repeats continuously and prevents the build up of dark current in the charge wells.

The number of row clocks sent at this stage, as well as the number used to read out the data described later in this section, is set to be the width of the CCD in pixels. In this case 388 pixels are being used per row but the hardware is able to handle CCDs with up 4096 horizontal pixels.

The clearing process will continue until the PC sends a signal that primes the circuit to allow a trigger input through.

When such an experiment start signal reaches the circuit, this clearing process is stopped and all the counters are reset to zero. The circuit waits for the falling edge of the experiment start and then begins issuing $10\mu s$ pulses to the column shifting circuits. These pulses are passed on to the three phase generators and then through the high current drivers to the CCD. They are also passed from the three phase generator into a counter which counts the number of full three phase cycles that have been sent. When a predetermined number have passed through, this stage of the operation stops.

The number of column clocks, i.e. the number of frames that can be taken, is slightly less than the vertical position of the slit as determined in section 3.2.3. This ensures that a few lines can be read out before the actual data, allowing these lines to act as a buffer area for charge build up in the CCD when it is being clocked down. If these lines were not present then this extra charge would corrupt the first few frames of useful

data.

The 10μs pulses are generated by a counter that counts to the value of $2^n + 1$. The most significant bit of the counter has a very long off time (counts from 0 to 2^n) and then a single count of high output before the system resets. This resetting is achieved with just one 2 input NAND gate in the usual way. As can be seen this method minimizes the column shift time, and thus allows the frame time to be reduced by using a different value of n. In this way 5, 2·5 or even 1·25μs frame times can be easily achieved. For the operation of this circuit, with an input clock of 12MHz, the value of n was chosen to be 7. This meant that the output pulses occurred every 129 cycles or once every 10·75μs. The minimum that would be easily achieved would hence be at n=1, corresponding to a frame shift of 4Mhz.

Theory dictates that the electron drift velocity within a CCD electrode field limits the maximum clocking speed of both the column and row electrodes. In the case of EEV devices analysis of this effect has shown that a normal TV sensor cannot be column clocked at greater than 4MHz without blurring of the data occurring, and hence the recommended running frequency is quoted as 2MHz. Because of this the clocks fed into the three phase cycle generator were set well below this at 0·7MHz (1·4μs periods) giving a full three phase cycle of 4·29μs.

When the photons hit the CCD surface they generate electrons that are pulled under the positive electrode. This will be the case for 6·46μs of the 10·75μs pulse. For the remaining part these electrons will be formed under a moving field and it is not clearly defined as to which electrode they will be attracted to. If only one position on the CCD surface is considered then, in normal conditions, it can be assumed that the positive electrode is in the centre of the slit and electrons will always be closer to this than to the surrounding positive electrodes which are three positions above and below.

If the situation where this positive electrode has moved down one position is considered next, then the electrons will only be nearer this if they fall into the bottom two thirds of the slit, otherwise they will be attracted to the upper electrode which is now just above the top of the slit.

For the next shift the electrons will now have to fall into the lower third to be attracted to the bottom electrode.

In this way it can be seen that it is well defined where the individual electrons will

be attracted but this position changes to the upper line at some point. As the shift is occurring for 40% of the time this period of uncertainty will be for one third of the clocking time or about 13% of the full 10·75μs period. Optimization of the clocking period would have to be carried out to improve this resolution.

The final stage is the read out of the data. The arrival of the experiment start signal triggers a switch to the slow row readout clock. This has no effect on the apparatus until the 10μs pulses have shifted the data down to the predetermined point. However, when this stage has completed the slow row readout clock is used to read out the data in a slow scan mode. This readout is done in a similar way to the CCD clearing stage before the experiment start pulse arrived. Rows are clocked out and at the end of an entire row a single column shift is automatically done to allow readout of the next row of data.

A full circuit diagram of the digital controller is given in inserted figure V.

3.3.2 Microprocessor generated controller signals

The original signals for this instrument came from an I/O card containing an 8255 driver. This allowed direct control from a PASCAL program on the PC and also had control of most of the signals including experiment start, fast or slow readout clocks and row readout enables. As the apparatus was developed further these were gradually automated with just a single line that acted as an enable switch when combined with the input signal available from the laser pulsing hardware. The I/O board also provided an 8 bit data bus and the required clocking signals that were used to read out the stored data.

3.3.3 CCD drivers and reset pulse generation

The final stage to be designed was the reset circuitry for the CCD. This has to be part of the output cycle and obviously the time between the data becoming valid and the reset pulse defines the maximum time available for the A/D converter to complete its operation. This reset pulse can occur at almost any time over the three phase cycle, but in this case it has been fixed so as to optimize the valid period of the output. Figure 3.5 illustrates the timing of this pulse along with the control signals sent to the camera head.

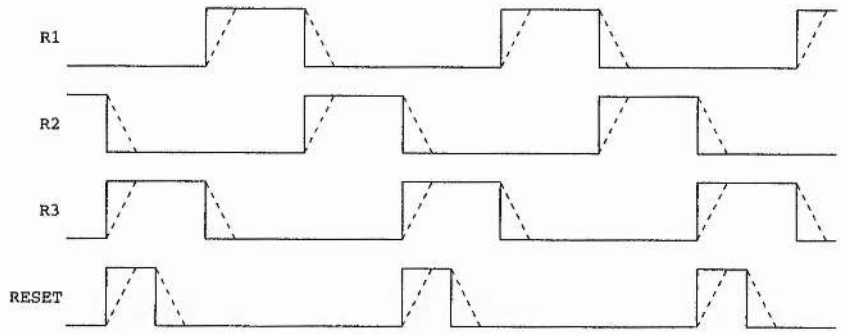


Figure 3.5: Reset timing for the CCD camera head.

The CCD requires 10-12V pulses to drive its row and column shift signals. These were generated from the DS0026 clock drivers. There are a number of different options to these, including the EEV hybrid parts and some Sony units. More details on such options are given in the appendix. Most of these are not designed to be used in non standard camera applications and were thus limited in their frequency ranges. Because of this the simpler DS0026 parts were used and were found to be very suitable to this type of drive application.

Because of possible crosstalk between the digital and analog lines the head was mounted remotely with the conversion to 12V being done on the digital controller board and then passed through ribbon cable to the analog head. It would probably have been advisable to also mount the voltage sources on this head to reduce the noise still further but, as will be seen later, crosstalk problems were so small as to become part of the background noise signal.

3.3.4 Trigger input pulse conditioning

The available laser driver, a Spectra Diode Laboratories SDL922, had a TTL output representing the current pulse used to trigger the laser pulse. However, when analysed it was found to produce a signal as shown in figure 3.6.

This clearly needed a conditioner to allow it to be used to drive the experiment start signal. The design of this stage was based around a standard comparator setup. Table 3.1 shows a list of available parts and the final circuit is shown in figure 3.7.

The LM311 was chosen for this application as it has a bandwidth just able to handle

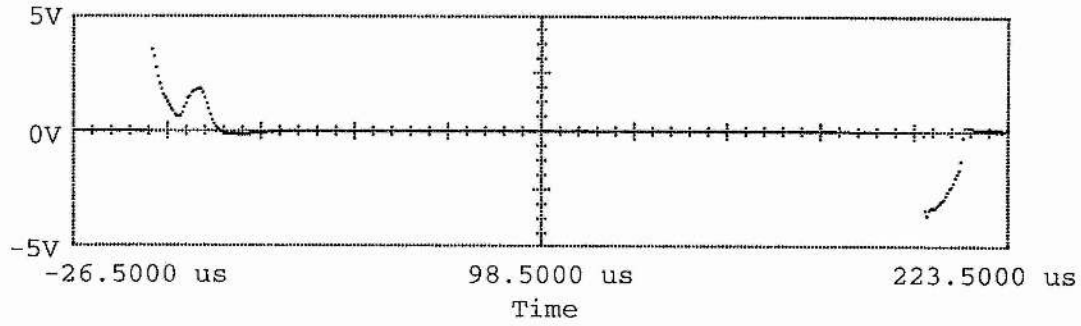


Figure 3.6: Oscilloscope trace of the laser pulse trigger output.

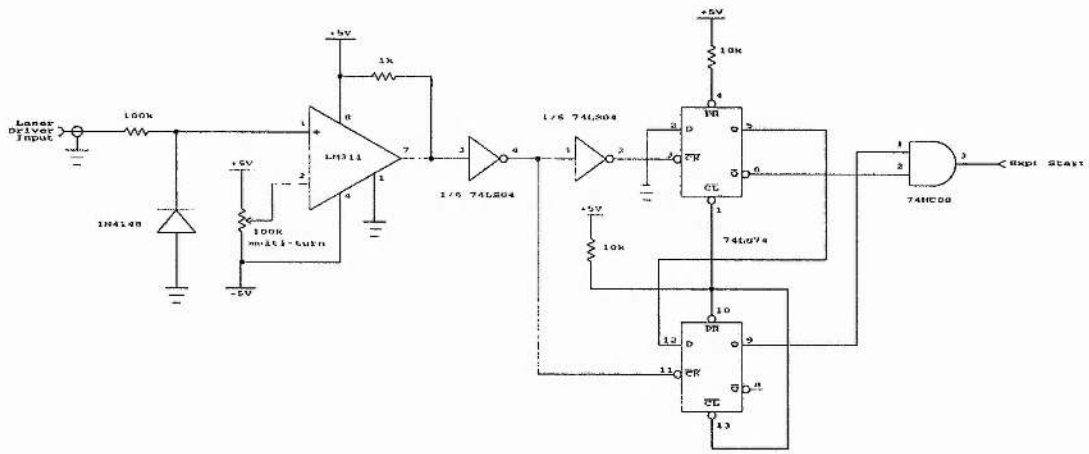


Figure 3.7: Circuit diagram of the input pulse shaper.

Device	Supplier	Comments
LM306	Farnell	Only TTL and OC output
LM311P	Farnell	Only 30V, TTL direct drive
LM311	R.S.	High current, high bandwidth

Table 3.1: A selection of comparator devices.

the required trigger speed.

A diode input limiter was needed to remove the negative going end pulse that would otherwise cause the comparator to saturate as the single supply rail output induces negative voltage mirroring of the input. The voltage setting was adjusted so as to make the unit trigger only on the primary start pulse, ignoring the secondary part. The input latches of the digital hardware required a minimum of $0{\cdot}5\mu s$ to latch the experiment start signal so the voltage setting had to be very precise so as to keep within the short pulse but also to provide a long enough pulse for the latches. This was achieved through a multi-turn pot.

3.3.5 IR laser analysis

Some of the more popular pulse sources are in the infra red spectrum, as generated by some solid state lasers. These have some advantages for this experiment as a fully opaque filter may be used to completely remove non IR wavelengths. This allows the camera to be run at normal room light levels without saturating the CCD.

The test laser used here, along with most medium length pulsed systems, has a minimum repetition rate of 10Hz. This means that manual resetting of the system could occur during a pulse and hence affect the results.

In practice it was found that the use of the PC to prime the system worked well with this repetition rate. In the event of faster system being used, then a shutter could be incorporated that used one laser pulse to trigger its opening, then stayed open for the actual pulse to be analysed and then closed for the next one.

3.4 The Digitizing Camera Head

3.4.1 The CCD bias voltage supply

The control inputs to the CCD consist mainly of the bias voltages and frame shifter signals. Because of the need for a self contained instrument it was decided that a dedicated power supply would be required to provide these bias voltages. The supply described in chapter 1 would have been suitable but it was felt that an alternative design for the low current supplies could be used to provide both a cheaper and simpler power source while still providing the functionality of the more complex designs shown in

figure 1.7.

A simple transistor stage was used to supply a constant voltage for the minor voltages (2V and 6V) and a 78L15 was used to generate a variable voltage for the 17V and 22V supplies. This was achieved by tying the ground inputs to a voltage divider and hence adding a fixed voltage level to the normal 15V output of this component.

The final circuit for this new power supply is shown in figure 3.8. This includes the 5V, 24V and -12V supplies which are required for both the digital circuitry and the output amplification and processing stages.

3.4.2 Analog signal conditioning

Processing of the CCD output stages is needed to first remove the DC offset ($\sim$ 17V), then remove the reset noise and finally amplify the signal to the required output level. For slow scan operations of the type used for this camera the usual practice is to use capacitive coupling to remove the DC component, then use a dual slope integrator to reduce the readout noise of the CCD output and finally use a simple amplifier to obtain the required levels.

In order to test this type of circuitry a dual slope integrator was built as in figure 3.9. This worked as predicted but suffered very badly from noise pick up in the stripboard used for its construction. It was noted at this point that the readout noise of the CCD was in fact lower than a single unit of digitization for an 8-bit A/D converter (i.e. less than the saturation point $\div$ 256). This meant that the converter could be driven from the direct signal without the need for reset noise removal provided the dummy output was used to correct for the reset level errors (see appendix).

Figure 3.10 shows the block diagram of the output processing stages. The first part was a simple differential amplifier using an NE532 dual op-amp package. This part is nominally specified to work at a maximum input offset voltage of 15V. It was found, however, that the amplifier still worked well at the 17V offset voltage directly available from the CCD outputs. It was worthwhile using the amplifier in this way as the ability to DC couple the signals meant a significant increase in the accuracy and linearity of the amplifier stages.

The next stages are a gain and offset adjuster for the A/D converter and then a limiter that provided active voltage clamping of the signal. This signal was then fed

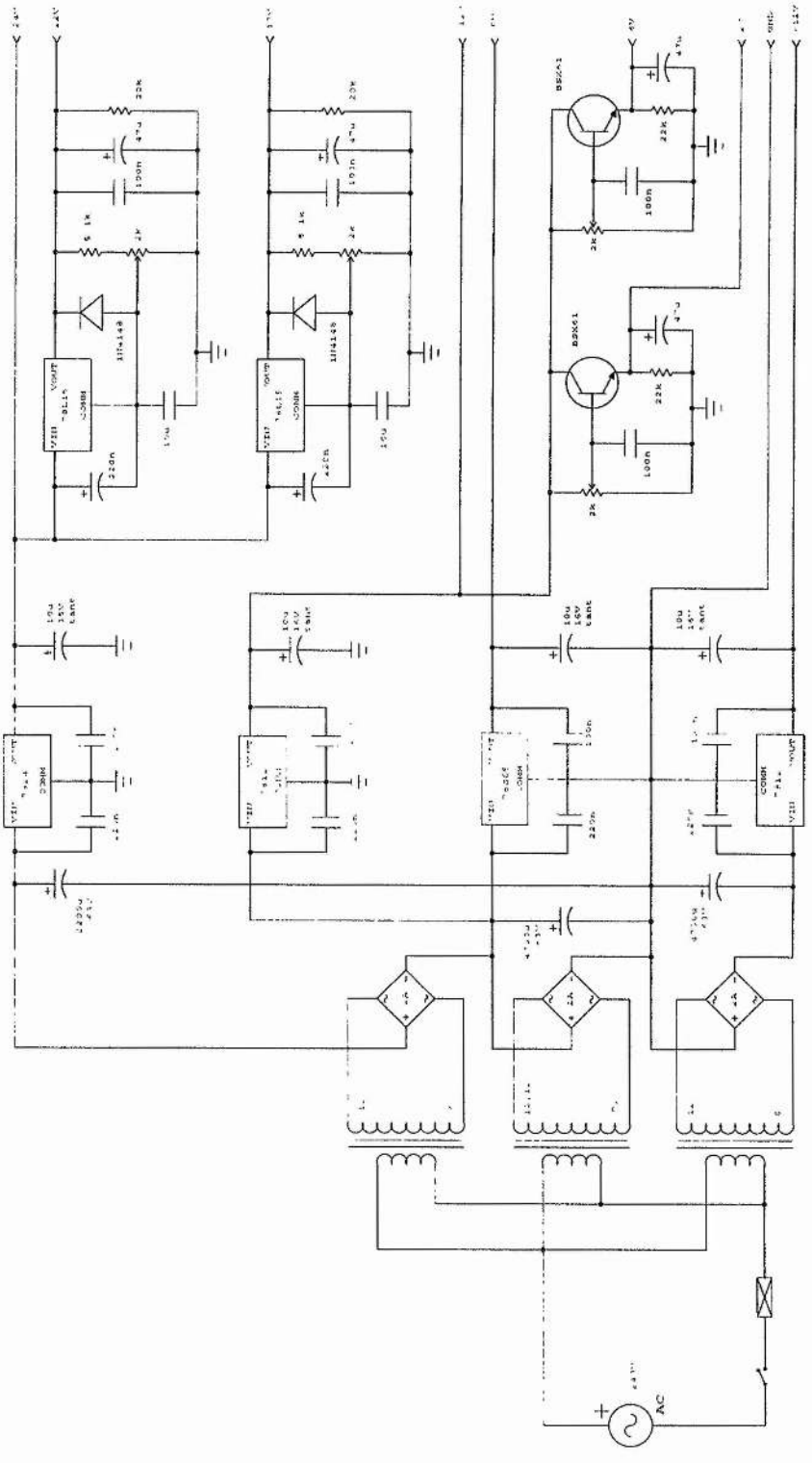


Figure 3.8: The power supply for the camera electronics.

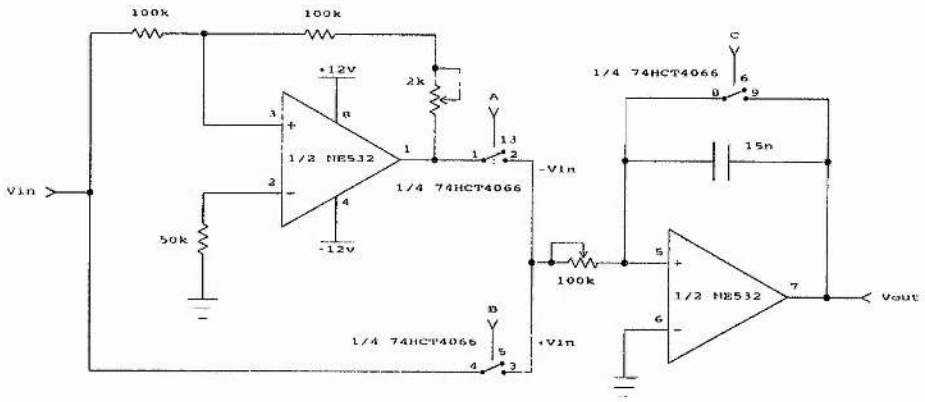


Figure 3.9: Experimental dual slope integrator.

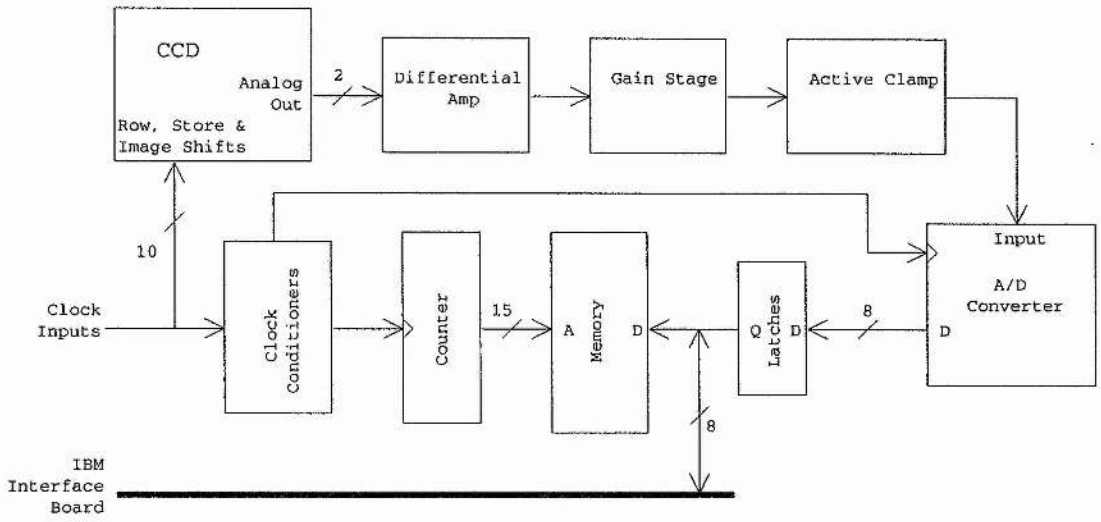


Figure 3.10: Block diagram of the camera head circuitry.

into the digitizing section.

3.4.3 The digitizer and data storage circuitry

Originally the A/D was the ZN439E. This is a successive approximation converter that has a clock period of down to 6μs and hence provided adequate conversion time when the systems was completely driven by a PC and the data was read out into the PC as it was digitized. Analysis of the results taken at this speed showed that the dark current build up was too large for this approach to provide suitable accuracy and so the digitizing stage was replaced with a faster converter, the ADC304, as used in chapter 1.

The A/D was set up so as to operate in the stable part of the output waveform. The reset pulses were optimized to allow this stable period to be as long as possible. The A/D timing was also set so that the shifting stages of the CCD were causing the least interference. This was achieved by using the fast clocks and shifting the phase of the A/D drive pulses so as not to coincide with the row clocks.

At the same time as the A/D was upgraded, a storage stage had to be added so that the higher conversion rates could be met without being limited by the PC communications rate.

Because this circuit was mounted remotely to the other digital sections, the system had to be self contained, taking a minimum of extra signals. Two counters were used, clocked from either the row readout pulses or from a processor control line. When the experiment start signal comes into this stage, the counters are reset and then, when the readout stage starts, they are allowed to increment. These provide the address signals for a 32k byte memory chip. The digital hardware supplies the correct timing for the row readout pulses enabling the data to be stored sequentially in this memory, one pixel readout cycle always corresponding to one memory location.

The nominal parameters for the hardware were to look at a period of 200μs in 10μs frames. This would require $388 \times 20 = 7760$ pixels. As the system can be run at much faster clock rates (up to 1μs per frame) this memory size would need to be correspondingly larger and the choice of 32k just about meets the maximum requirement.

When the full 32k memory has been filled, this part of the circuit latches up,

preventing further writes to the memory until the experiment start signal resets the circuit. This latching also passes onto the memory $\overline{RD}$ and $\overline{WR}$ connections, switching them so that the memory now puts its addressed data onto the PC interface lines rather than reading data from the A/D lines.

The processor on the PC downloads the data by toggling a direct control line into these address counters. This causes the counters to clock back to zero and hence present data to the 8 digital lines. Further pulses on the control line sequentially move through the address space, the PC simply reading the data bus and then clocking on to the next value.

A full circuit diagram of the readout and data storage circuit can be found in inserted figure VI.

3.4.4 Finding the noise levels on this detector

In order to find the noise of the amplification stages it was necessary to remove the CCD and replace its outputs with a known, constant, voltage level. This meant that the loading on the row shift lines was not as it would have been were the CCD to be left in the circuit. As the digitization points were chosen so as to act when no shifts were occurring, this was not as large a problem as it could have been.

The substituted voltage level was set to about half the CCD saturation voltage level, about 100mV, and an experiment run commenced. The mean output would now be the constant voltage source and the variance of this would be a measure of the noise levels. This assumes the noise level of the reference voltage source is low when compared to the circuit noise being tested. Given the source was a fully stabilized, regulated power supply, this was felt to be a reasonable assumption.

In order to analyse the results a program was written to examine the data values and find their distribution. The noise was, predictably, Gaussian in nature and over a sample set of 388×80 data points had a standard deviation of 0·941 ADUs.

The A/D used here has an inherent error of $\pm 0{\cdot}5$ unit. This is an absolute error, always causing the result to be rounded by a random amount between 0.5 and -0.5. The cause of this error is in the way that most flash analog to digital converters operate. They choose an output value by comparing the input voltage to a set of known, fixed voltage levels. Internally there is a set of comparators, each of which is fed with one

of these fixed voltages and the input voltage. For any given input all comparators fed with a fixed voltage level that is less than the input voltage will turn on. The comparator outputs are then encoded to a binary value and provided as output. The effect of this is to make the A/D round down any given input to the lower signal. To improve this the voltages that are fed to the comparators are compensated to give a monotonic reading at 0 and 255, and hence the rounding is $\pm 0{\cdot}5$ rather the between 0 to +1.

Because of this behavior it is not easy to analyse the A/D error in isolation. One possible way to do it would be to treat the A/D error as a normal, Gaussian, random noise source, with $\sigma = 0{\cdot}5$. This would ensure error calculations were always within the correct parameters.

Fortunately a lot of data is available and this allows the effect of the A/D and the amplifier noise to be considered as one noise source acting as a dithering value to help calculate the deviations. The value given above is thus representative of the compound error without the need to look at the causes separately.

This value is easily good enough for the purposes of the camera and, with the techniques for dark current subtraction described in subsequent sections, provides a very stable starting point for data analysis.

3.4.5 Bandwidth considerations

The bandwidth of the amplifiers was a major concern in this design as the high speed converter being used required a stable input to produce a reliable conversion. Tests of this were carried out by passing a square wave into the amplifier system and then analysing the output. The amplifiers were found to be slightly over-damped but had a full scale bandwidth of twice the pixel rate. This meant that it was possible to stabilize the output from a minimum to a maximum change on a pixel boundary and verified that the design bandwidth was within the required limits.

3.4.6 Correcting for gain distortion

A major problem with all digitizers is the fact that they are usually not perfectly linear and are also not perfectly uniform in the voltage range that each unit converts over. Because these errors are systematic rather than random in nature they can be corrected

for.

Having built the amplifier and digitizing stages and verified the noise levels as being small, the next requirement was to correct for distortions both within the digitizer and within the amplifier stages driving the converter. To do this the amplifiers were calibrated to provide the correct gain and DC offset for normal operation. The CCD was then removed and the amplifier input pins were connected to a triangle wave generator.

The signal generator was adjusted so as to just saturate the input at both the minimum and maximum voltages. When triggered the camera now acted as a simple digitizing oscilloscope, providing many frames showing the triangular waveform. The waveform could now be used as an indicator of the full scale gain of the amplifiers and allowed the DC offset to be analysed.

The amplifier gain was set to give a full scale reading of slightly larger than the maximum signal out of the CCD to ensure the complete range was covered. This was set up by trial and error using a helium-neon laser source. Once the signal generator was adjusted to give a full scale output the signal was found to be approximately 280mV. This compares well with the 200mV lower bound for the maximum CCD output as specified in the literature.

The DC offset was set to 20mV, or so, so as to allow the minimum output levels to be clearly visible in the results.

The final stage in the analysis of the amplifiers is in the linearity of the A/D itself. The ADC304 converter has an input pin linked to the centre tap on the converter resistor chain. This can be used to provide better linearity adjustment by tying this pin to a voltage reference set half way through the conversion range. This is set incorrectly in figure 3.11 and illustrates the distortion correction needed for the triangle wave input.

A program was written that took the data set and extracted the waveforms by looking for the first non zero value, reading till a saturation value was detected and then repeating the process. Because of the complexity of such an algorithm each line was output to the user to allow false readings to be discarded manually. These waveforms were then averaged in order to remove background noise in both the instrument and the signal source. The result of this should have been a straight line were there no distortion.

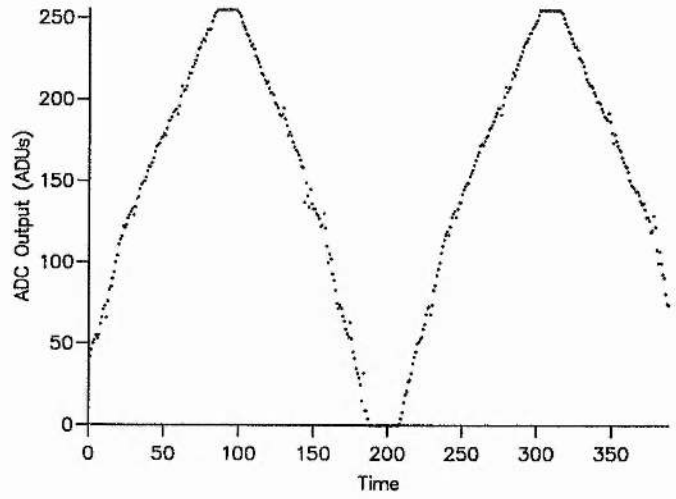


Figure 3.11: A triangular waveform used for distortion calibration.

By finding the equation of a line passing through both 0 and 255 this average data set enabled a correction mapping (in the form of a lookup table) to be deduced and from this the required correction factors for each pixel could be found. Each actual conversion reading is simply used as an index into the lookup array to find the correction factor. Figure 3.12 shows the results of this where the input (x-axis) is mapped to a corresponding output (y-axis). As the resolution of the output mapping is higher than the input resolution the resultant output has to be non-monotonic. The values of 0 to 1000 for the output range gave a suitable resolution with only a small increase in required processing time. Thus the results can now be corrected for all gain distortions between the CCD output and the end of the digitizing process.

3.4.7 Initial calibration

Having got readings for the voltage input against A/D reading, it is now possible to convert to charge and hence number of electrons detected. As only relative effects are needed for this application this is not too important but, if needed, can be deduced from the data given in table 2.1 and the subsequent calculations.

From this data it is possible to show that an output range of 200mV corresponds to 339×10^3 electrons. As the full scale output for this hardware is 280mV, then each

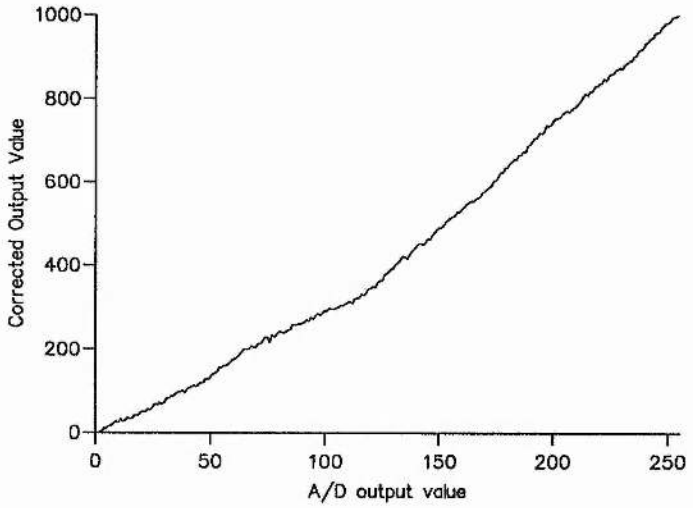


Figure 3.12: The distortion correction mapping.

division corresponds to

$$\frac{1}{255} \times \frac{280}{200} \times 339 \times 10^3 = 1861e^- \text{ per ADU.}$$

It was also possible to look at the saturation point for the sensor. This was done by increasing neutral density filtering of a HeNe laser until the sensor output was half way through the full range. This was a good working point, given the dark current build up that will be discussed later.

Once this half saturation point had been found, the laser power was measured and the beam profile was captured. The energy falling onto the CCD is now some fraction of the full beam energy, as measured by the power meter. Assuming the laser profile is Gaussian, it is possible to work out the peak intensity by considering the two dimensional Normal curve equation

$$f(x,y) = \frac{P}{2\pi\sigma^2} e^{-\frac{1}{2}\left(\frac{x^2+y^2}{\sigma^2}\right)}. \tag{3.1}$$

Where P is the total power of the beam (in Watts). For a HeNe laser of 0·806mW output power the measured beam power, after passing through a set of neutral density filters, was $0{\cdot}343 \pm 0{\cdot}001\mu$W with a background power measured at $0{\cdot}024 \pm 0{\cdot}001\mu$W. This results in a power to the slit of $0{\cdot}319 \pm 0{\cdot}002\mu$W when the room is completely dark.

By looking at a plot of the measured pulse as output by the experiment, it was seen that the peak intensity was 141 ADUs and hence the half peak height was 70·5 ADUs. The half peak height was measured at a point approximately 8 pixels from the peak, i.e. $8 \times 22{\cdot}5\mu$m or 180μm.

For rigorous treatment, the required analysis is to look at the double integral

$$\int_{-\delta}^{+\delta}\int_{-\delta}^{+\delta} f(x,y)\,dx\,dy.$$

However, as δ (half the pixel width) is small compared to the size of the Normal, it is possible to approximate this to simply the area multiplied by the centre height of the function.

This means that, by taking the ratio of the maximum peak height to the half peak height the calculation to find σ reduces to

$$\frac{f(x,0)}{f(0,0)} = e^{-\frac{1}{2}\frac{x^2}{\sigma^2}} = 0{\cdot}5$$

and hence, with the x separation being 180μm, σ is found to be $1{\cdot}53{\times}10^{-4}$.

By substituting this into equation 3.1 it can be seen that the peak value of $f(x,y)$ occurring at $f(0,0)$ is

$$\frac{P}{2\pi\sigma^2} = \frac{0{\cdot}319 \times 10^{-6}}{2\pi(1{\cdot}53 \times 10^{-4})^2} = 2{\cdot}172 \text{ Watts m}^{-2}.$$

3.5 Dark Current

3.5.1 Theory of dark current build up

Dark current, otherwise known as the 'fat zero', is the build up of thermally excited electrons in the charge wells of the CCD [70]. When the electrons in the semiconductor get excited into the conduction band, near to one of the CCDs conductor electrodes, they are trapped by the potential under that electrode and appear as signal when read out.

Clearly this effect is dependent on both temperature and time. The build up with time is generally believed to be linear, but for large charge levels some saturation effects may be seen [71]. These are caused when impurities in the silicon absorb some of the free electrons. Once these have all been saturated the gain of the silicon should suddenly increase. As can be seen from results given later this effect is too small to

notice and the subtraction of dark current will take care of a large part of this problem anyway.

The dark current build up with temperature arises from the increased probability of an electron achieving the higher energy state when it has higher thermal stimulation. This follows the diode law

$$I_D = Ae^{\left[\frac{-V_{BG}}{\left(\frac{2kT}{q}\right)}\right]}. \tag{3.2}$$

Under normal operating conditions V_{BG} is 1·1 volts. Using this, it can be seen that a 1K drop in temperature will correspond to as much as a 3.5% change in dark current. However, over the sort of periods being considered here (about 20 seconds for a single experiment), ambiant temperature was very stable and this amount of dependency did not cause any problems.

For the purposes of this experimental work, the sensor was not temperature stabilized, which would have required the use of a cold finger or Peltier system and the added complexity of having to prevent condensation via techniques such as using a vacuum envelope or passing nitrogen gas across the sensor surface. The added size and extreme mechanical complexity of these options meant that software correction would be the preferable solution. Because of the large temperature dependence of the dark current build up it would be necessary to correct the data at the time of its collection rather than using some fixed correction method for all data sets.

Some interesting work is being done into very low dark current devices [72], they are not yet commercially available however, and it may be some years before significant reductions in dark current levels are seen.

3.5.2 Dark current build up during a sequence of frames

When run in complete darkness for 80 frames, the slow scan method of data collection produced a set of blank frames that illustrated the build up of dark current. Figure 3.13 shows the average reading per line, plotted against the line number. This is in effect a plot of dark current reading against time. This graph shows that the data can be successfully collected provided the illumination, when present, is kept to about half of the saturation level, thus allowing the dark current to build up without saturating the sensor. From the results of section 3.4.7 and a best fit line for figure 3.13 this can

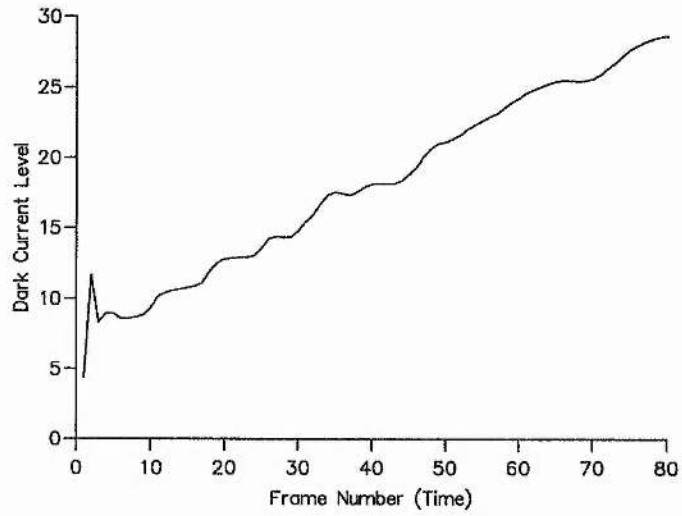


Figure 3.13: Average dark current per line vs. line number.

be seen to correspond to a dark current build up, in one cell, of

$$\frac{20 \times 1861}{70 \times 388 \times 85{\cdot}3\mu s} = 16066 e^{-} s^{-1}.$$

The quoted value for this sensor [4] is 20000 electrons per element per second, so this compares favorably.

3.5.3 The repeatability of dark current readings

Possibly the best solution to correcting for dark current is to be able to subtract a set of readings with no light (i.e. just dark current), from the one with both dark current and the required signal. The resultant should be the required signal but with added noise due to the error in the dark current readings when they are subtracted.

To verify that this technique can be used it must first be established that the error in the dark current readings is small enough to produce acceptable results.

To do this a set of five experiments were performed in close succession in total darkness. These were then compared via computer processing.

In order to fully analyse the data a best estimator for the error has to be found. Given that the dark current is discrete it should be possible to approximate its Poisson distribution with a Normal distribution, $\mathcal{N}(n, n)$. The standard deviation for thermal build up is then dependent on the square root of the number of excited electrons. Given

the data in chapter 2, table 2.1, and using this Normal approximation, it is possible to show that the deviation of the dark current when it reaches half the saturation level, (an A/D reading of 128) is $\sqrt{\frac{339\times10^3}{2}}$=411·7 electrons, and hence the error is about $\frac{1}{3}$ of an ADU.

This is small, and visual analysis of the data confirms this slightly increases the deviations of readings at large dark current values.

The other effect on the deviation is the dependence of dark current with temperature. This should be dependent on absolute temperature change as defined in equation 3.2. The slight changes in temperature over the few seconds of data collection, will have an approximate linear effect on the deviation build up. This should be minimal as the fractional change in absolute temperature is at most $\frac{0.1}{293} = 3{\cdot}1 \times 10^{-4}$.

As an approximation to the overall error for the experiment, rather than use a fractional error, the mean variance for the full data set was identified as a suitable best estimator.

This can be calculated by finding the mean of all the variances for each point in the five data sets. Any one variance, being based on only 5 samples, will be very inaccurate, but the mean over all 80×388 points will be much better defined.

The variance for this is defined by

$$\mathrm{Var} = \frac{\sum_{j=1}^{N\times M}\left[\sum_{i=1}^{P}\frac{(x_{i,j}-\overline{x_i})^2}{P-1}\right]}{M\times N}, \tag{3.3}$$

where N is the number of lines in the experiment, M is the number of data samples in one line and P is the number of full data sets being averaged over.

In order to reduce this to a computable form, it is necessary to be able to build up intermediate results as analysis progresses through the data. The well known reduction of the standard deviation formula gives a new form of equation 3.3 defining

$$\mathrm{Var} = \frac{\sum_{N,M,P} x^2 - \frac{1}{P}\sum_{M,N}\left(\sum_{P} x\right)^2}{(P-1)\times M\times N}. \tag{3.4}$$

This can be calculated by keeping a set of values representing the $\sum x_i$ values, one for each of the P sets. As this is a small number (5 in this case) the result can be computed without too much difficulty.

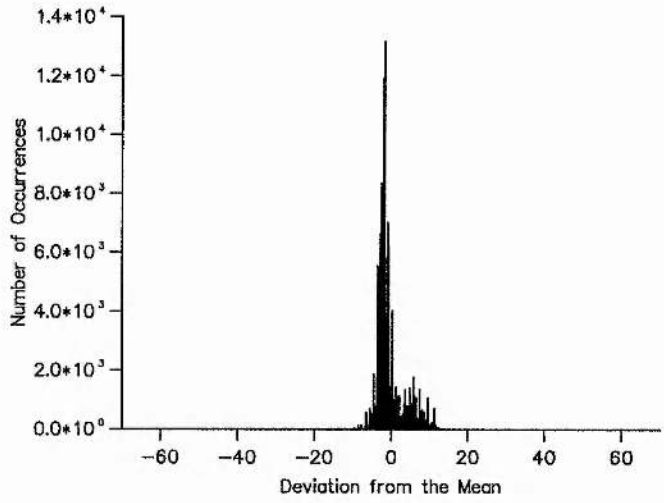


Figure 3.14: The distribution of the deviations for 5 dark frames.

The only other limiting feature of this method is that it requires, in this case, the storage of numbers as large as the sum of $(388 \times 80 \times 5)^2 = 155200^2$ (of order 10^{11}). This is potentially larger than the maximum integer capability of most programming languages. In reality, this was not found to be too serious a problem as most of the numbers actually involved are smaller than the maximum 255 output and hence reduced the calculation to lie within the programming language constraints.

When processed in this way the data showed that the reproducibility of dark current was very good. Using the standard deviation derived from $\sqrt{\text{Var}}$, and using equation 3.4 to process the data, the deviation was found to be 1·352 corrected ADUs. Observation of the data in the worst case showed that the error can reach ± 10 corrected ADUs, as seen in figure 3.14 which shows a histogram built up from the deviations of every point from the mean of 5 corresponding points within the 5 data sets. It shows the noise to be very small compared to the peak output and also shows that the noise does have an approximately Gaussian spread.

This value includes the noise of the amplifier section as well. The result found in section 3.4.4 can be subtracted from the above noise value to obtain some measure of the reproducibility of the dark current that is independent of instrument noise.

If two independent Normal distributions N_1 and N_2 are subtracted then the resul-

tant distribution has a standard deviation given by

$$\sigma_{N_{tot}} = \sqrt{\sigma_{N_1}^2 + \sigma_{N_2}^2}.$$

From this the resultant error in the apparatus, σ, was shown to be 0·97 corrected ADUs for the mean noise value, and approximately 11 corrected ADUs for the worst case value.

These noise levels are small compared with the range of the converter and hence the use of software error correction will give the required accuracy.

In order to deal with the errors rigorously, it is necessary to look at the form of these noise values. For the instrument noise the distribution is known to be approximately Gaussian and hence it can be expressed as a standard deviation, σ_i.

For thse dark current reference frames, the noise figure, ζ_d, will be of the form

$$\zeta_d = \sqrt{\sigma_i^2 + \mathcal{G}^2(n)} \tag{3.5}$$

where $\mathcal{G}(n)$ is a function of the number of electrons, n, and involves both the linear temperature dependence ($\sim n$) and the Poisson distributed dependence ($\sim \sqrt{n}$).

3.5.4 Subtracting 'dark' frames from a frame sequence

Having verified that the dark current is reproducible within stable parameters it is now possible to subtract it from a data set provided that the time between taking the dark frames and the actual results is small, thus guaranteeing similar temperatures. For even higher accuracy it would be possible to take a few dark frames before and after the experiment and then average over these. As the temperature in the available laboratory facilities only changes noticeably over a period of hours, this was felt to be an unnecessary complication but in extreme cases could be a useful technique.

Given the error, ζ_d in a single data set, the mean of 5 such data sets will be known to $\frac{\zeta_d}{\sqrt{5}}$. Hence, when the mean dark frames are subtracted from the result frames, the new noise figure, ζ_r, will be given as

$$\begin{aligned} \zeta_r &= \sqrt{\zeta_d^2 + \frac{1}{5}\zeta_d^2}, \\ &= \sqrt{\frac{6}{5}}\zeta_d. \end{aligned}$$

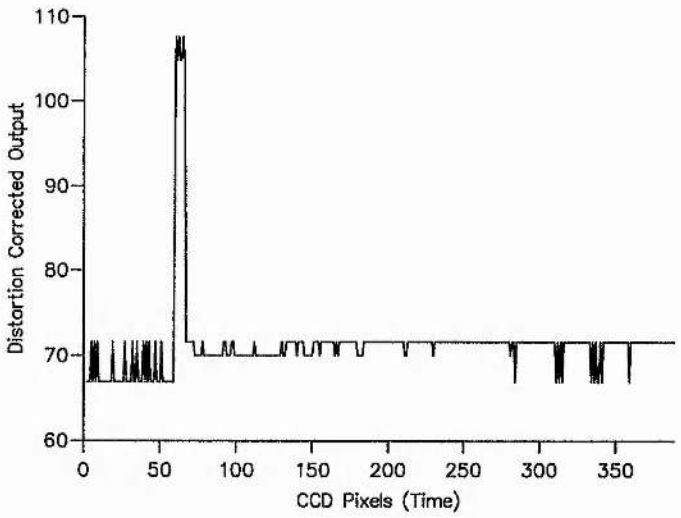


Figure 3.15: A single line output from a dark frame.

By substituting in equation 3.5 it can be seen that the rigorous treatment leads to

$$\zeta_r = \sqrt{\frac{6}{5}}\left(\sqrt{\sigma_i^2 + \mathcal{G}^2(n)}\right).$$

Using the mean figure and maximum values for $\mathcal{G}^2(n)$, the results are found to have an error figure of 1·48 corrected ADUs and 12 corrected ADUs respectively.

3.5.5 Finding pixel widths from dark current readings

Figure 3.15 shows a single output dark line. As can be seen there are a few pixels that appear to have a much greater signal than others. This is due to the fact that dark current is also dependent on the area of silicon involved [73]. Though CCDs are made to close tolerances, pixels are not always the same width, though it is true to say that all pixels in a single column have the same width. This fact can be used in conjunction with the above dark current readings to analyse the surface of the CCD and correct for pixel size variations.

The mean frame of dark current readings allows a very accurate record of the actual dark current build up parameters. As the data is only shifted between rows when it is read out, almost all the build up will occur when pixels are being stored somewhere within the CCD. Once the time dependency of the build up has been established for the entire experiment, by fitting a line to the entire data, it is possible to predict the

values for each single pixel with some accuracy. The difference between this predicted value and the actual reading for each pixel is an indication of the relative size of the pixel. Since the average pixel size is well defined at 22·5μm, this difference is also a measure of the physical size of the pixels.

If this process is used for the entire frame, then many rows of data will produce values and these can then be averaged to produce even more accurate values for the pixel widths.

In theory, the same technique could be applied to looking at successive rows and hence a pixel size map can be built up for the entire CCD surface. In reality the pixels were found to be too small in variation for this last phase to be detected within the range of the available equipment.

3.6 Results

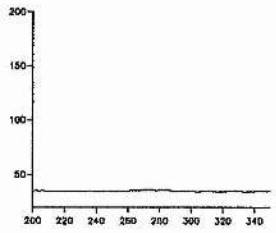
3.6.1 Fully corrected frames from a pulsed laser

The test laser for this apparatus was a diode pumped solid state laser producing 150mJ at 810nm and 1064nm. This is based around a NdYAG and AlGaAs crystal laser which produces a wide range of pulse shapes covering about 200μs maximum. These lasers can be set to trigger in different ways depending on the current applied and the pulse rate. In this case the laser was set to produce a fast peak followed by a couple of smaller peaks, as Gaussian profile pulses.

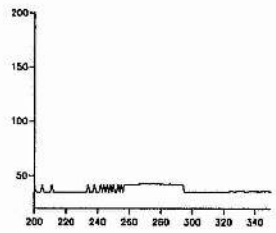
Having set the camera up to examine the output from this laser a few frames were taken and processed by removing the dark current noise as described earlier. Some of these frames are illustrated in figure 3.16. These illustrate the build up and decay of the primary laser pulse. The results were not as good as hoped for but still proved the principles of operation. The problem, as can be seen from the example frames is that the CCD is smearing the data out.

Two possible causes were identified, either the coupling was not good enough and the light was being spread out over a number of lines or the frame shifting was not very efficient and the trapped charge was being smeared across the surface of the CCD as it was being read out.

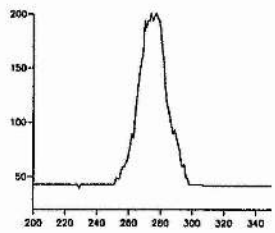
Tests using a beam chopper showed that the latter was not likely to be the cause



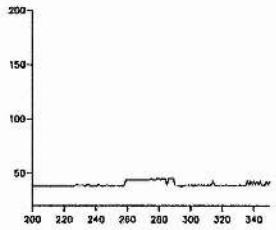
(a) Frame 9.



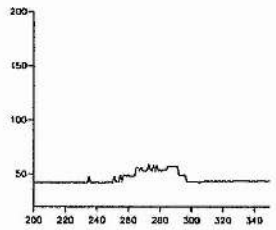
(b) Frame 12.



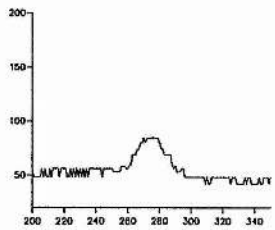
(c) Frame 15.



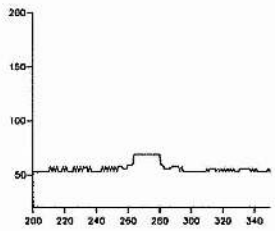
(d) Frame 18.



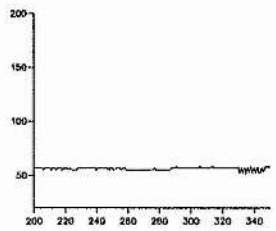
(e) Frame 24.



(f) Frame 27.



(g) Frame 33.



(h) Frame 36.

Figure 3.16: A selection of lines though a laser pulse.

as pulses generated by the beam chopper seemed to be slow on the rise as well as the decay side. If frame smear was to blame then only the decay side of the pulse should be affected.

This being the case the solution had to lie within the coupling mechanism. Unfortunately there was very little that could be done to improve this without getting some vendor redesign. Consultation with EEV showed that the mask layer, normally used to shield the store section, could be redesigned to allow a perfectly aligned slit to be made actually on the silicon surface. It was not believed possible to mount any external slit onto the CCD surface as the sensitive silicon requires chemical means to adhere things to it and glues would simply destroy the sensitive surface.

Another coupling problem would be if the beam is saturating the sensor. If the intensity was too great then the trapped charge would spread out over the surrounding pixels. This possibility was investigated by mounting the slit on a normal camera system. As the light could be contained within the one line, this was deemed to be a valid test. Calculations could then be made to produce a similar power into the slit with the unusual shift times being taken into account. It was also verified that if a sharp edge was mounted cutting off part of the beam near to the slit, then this produced a similarly sharp cutoff for the light seen in any one line, thus ensuring that the horizontal spread was not a degrading factor in the image resolution.

It was eventually decided that a combination of the slit mounting and the frame shift was the cause of the problem. The effect was to reduce the frame time to only about 50μs. at which point the spreading out effect could be virtually eliminated.

Future analysis would take two routes. Either the column clock dependency could be analysed and maybe increased to the 3MHz possible from the existing hardware, or a very precisely aligned 6μm slit could be used. If the alignment could be made to within $\frac{1}{3}$ of a cell (i.e. over just one electrode), then the smearing of electrons would be almost completely eliminated. It should also be possible to use some form of adhesive to mount the slit very tightly to the fibre faceplate thus reducing light leakage.

3.6.2 Displaying image sequences as surfaces

One of the biggest problems with displaying frames from a sequence is seeing the detail in a usable from. The best method for analysis is a video type picture where the pulse

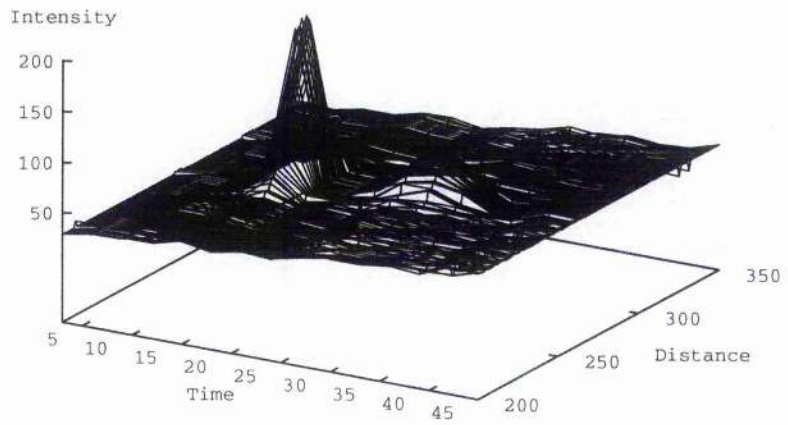


Figure 3.17: An example data set plotted as a (*distance, time, intensity*) graph.

is seen to run through its phases in a couple of seconds. The overall structure can be seen and if controls are added to allow individual frames to be seen as with a pause function on a VCR, then the detail can be further analysed.

One other useful form is to plot a 3D graph of time against distance across the pulse against intensity, i.e. a surface plot of the data. This form allows an overall view of the image to be presented and also provides the ability to see effects progressing through the pulse that are not so easily seen when individual frames are viewed without their relationship to each other.

In chapter 4 both software and hardware solutions to this type of display will be covered. Figure 3.17 shows the test data plotted as a surface representation. The pulse shapes can be seen and their relative timing is also made obvious.

3.6.3 The power vs. time relationship

Figure 3.18 shows an example plot of output power against time for one form of the pulse setup, found via a digitizing oscilloscope and a fast optical diode. It should

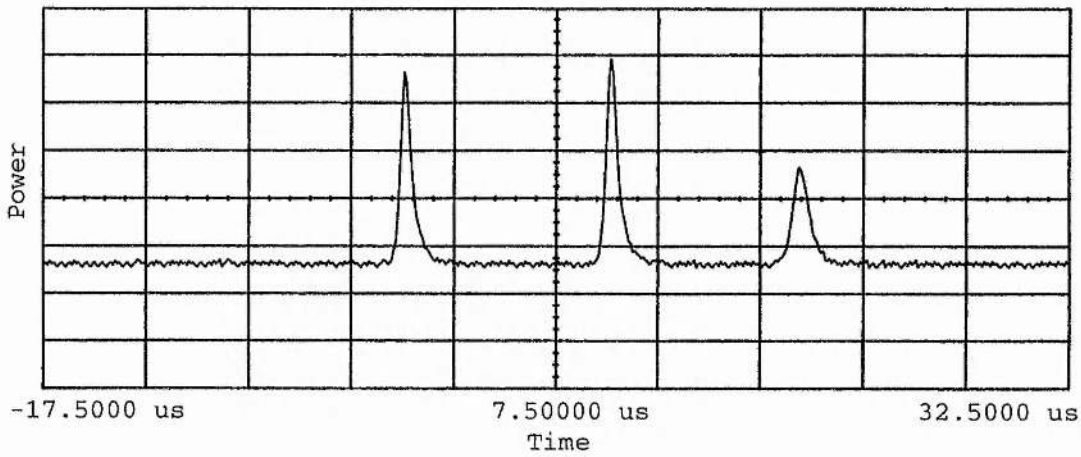


Figure 3.18: Power against time for a NdYAG laser pulse

be noted that the pulse positions are the same for those used in figure 3.19 but the intensities of the peaks varies greatly for pulse to pulse.

In order to test the camera with a comparison, a program was written to integrate the power in the pulse frames and hence produce a power against time relationship from the camera data. This program summed all the data values for a line and used that value as the power rating for this line, effectively integrating the beam power and hence representative of the pulse at that time.

The results of this are shown in figure 3.19. As can be seen the power spread across the surface has had the effect of smearing the pulses out along the time axis. Even so it is still possible to see the basic structure.

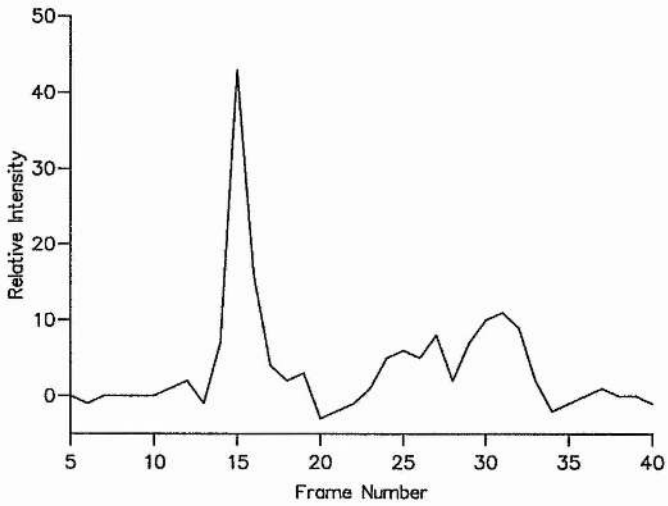


Figure 3.19: Power against time generated by frame integration.

Chapter 4

A hardware Solution to Surface Generation

4.1 Introduction

4.1.1 Objectives

The intensity of light emanating from an area of a surface is affected by many factors, the most obvious being the intensity of the incident light and the reflectivity of the surface. The study of representing different objects has become a vast research area and a large amount of work has been done on generating faster and more accurate models for computer imaging. The most common form of hardware solution has been to carry out computer analysis of the image data as fast as possible. While this has led to some of the spectacular images seen in modern computer literature it has not yet addressed one vital area of object representation - namely the representation of intensity surfaces on two dimensional screens in real time. It is one of the purposes of this chapter to remedy this situation.

The simplest form of this type of data is that generated by area detectors where a sensitive area, in the (x, y) plane, is used to detect an intensity signal at regular spacings of x and y within that area. Since CCD image detectors produce $(x, y, intensity)$ data values representing the intensity of incidental light on a surface they form an ideal source of such images.

As described in chapter 3 the need to display such 3D data on a 2D screen has not been adequately explored. In particular the requirement for fast, real time displays has become apparent as complex rendering algorithms take time and require expensive computer analysis hardware. One solution to the display of area detector images will be

presented here along with an investigation into the sort of hidden line hardware needed to provide intelligible representation of these intensity surfaces. A few improvements to the standard algorithms that have been used for simple image representation will also be discussed

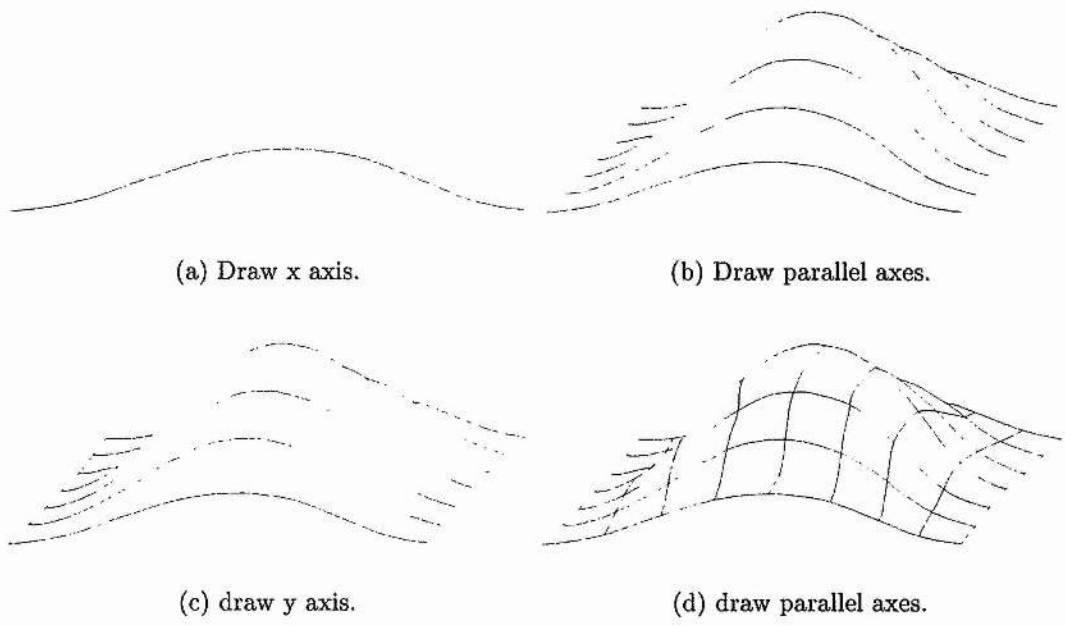
4.1.2 Full wire frame generation

The use of wire frame representations allows graphical data to be viewed in such a way that the gradients on the curves are visible. Lines that are close together represent shallow gradients while those that are widely spaced indicate large gradients (the opposite of contours on a map).

Another major advantage of surface, rather than grey-scale, image representation is that details are shown as linear effect. Unfortunately both the human eye and most graphics displays have non linear intensity responses to grey-scale images that prevent them from being able to differentiate between small changes in very bright or very dark images. Chapter 1 shows these effects along with some possible image enhancements that improve the information content for the human eye. Examples of the visual benefits of surface representation will be given later on in this chapter.

Figure 4.1 shows how a wire frame surface is displayed on the screen. Two adjacent points in the real world grid are mapped onto the screen, and then the two screen points are joined by a straight line. Lines are drawn so as to connect all adjacent points along the x-axis (figure 4.1a) of the image and then along some lines parallel to this (figure 4.1b). Next, lines are drawn connecting all the adjacent points in the y-axis (figure 4.1c) and finally some lines parallel to this axis (figure 4.1d).

The representation of the surface also depends on the observation point relative to the surface. For example, if the intensity is constant for all points on the surface, and the surface is viewed from a long distance down the z-axis (*intensity*) a simple grid would be displayed on the screen (figure 4.2a). If the viewing point is then rotated round the x-axis, the displayed image would change so that the squares compress in one direction (figure 4.2b). This eventually becomes a single line when the image is viewed from along the y-axis i.e. viewed edge on. If the viewing position were to be rotated about both the x and z axes then the plane would be viewed as in figure 4.2c. This is the type of view that will be dealt with in this chapter. The correct mapping



(a) Draw x axis.

(b) Draw parallel axes.

(c) draw y axis.

(d) draw parallel axes.

Figure 4.1: Building up a surface representation.

from world to screen coordinates depends only on the rotation angles to the x and z axes and the (x, y) coordinates in the image surface.

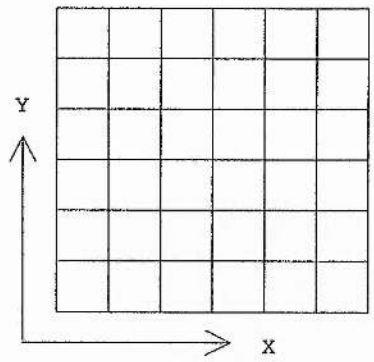
Figure 4.3 shows how the surface is viewed (in only one dimension). It also illustrates how the perspective is kept correct by this method of displaying the data. The value of h_x can be found via

$$\begin{aligned} \sin(\theta_1) &= \frac{h}{d+x}, \\ &= \frac{h_x}{d}, \end{aligned}$$

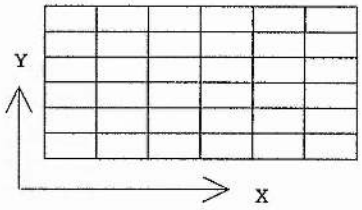
and hence

$$\Rightarrow \quad h_x = \frac{h \times d}{d+x}.$$

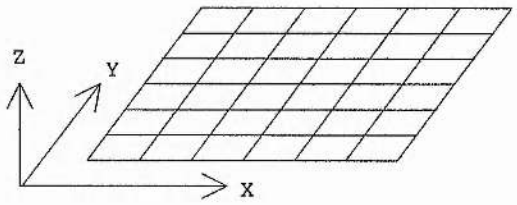
Clearly, the height of the image is proportional to the inverse of the distance between the eye and the object. This theory can be extended to find h_y (3D to 2D) and even h_z (4D to 3D) representations



(a) Viewing from above.



(b) Rotated about the x axis.



(c) Rotated about both x and y axes.

Figure 4.2: Changing the viewing position of a plane.

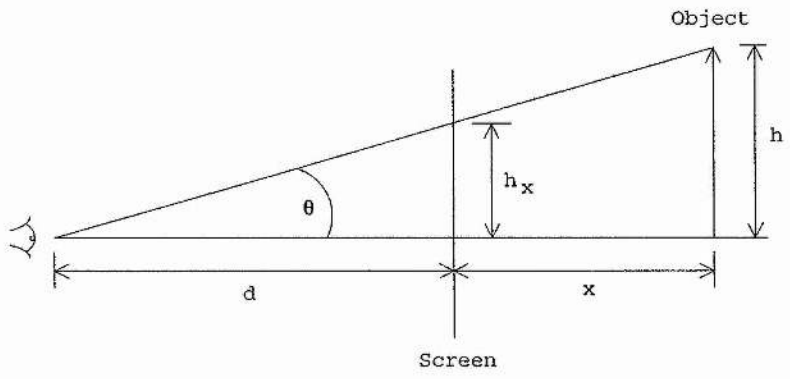


Figure 4.3: Viewing an object through a screen

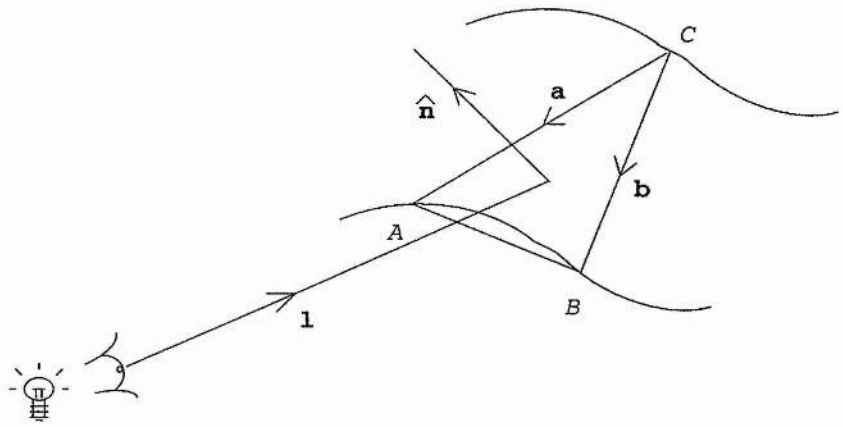


Figure 4.4: The vectors representing a small surface area.

4.1.3 Ray-tracing and shading

Shading is one of the more complex parts of surface generation. The usual simplification for ray tracing algorithms is to remove texture calculations when displaying a surface. This is justifiable since, as there is no physical reality to an intensity surface, an algorithm must create some form of artificial texture for the image. The second simplification is to ignore reflections between some parts of the surface and others. Once the coordinates of the light source have been specified, it is simply necessary to calculate the angle between a vector from the light source to the object and a vector normal to each surface element. The cosine of this angle represents the reflected light fraction that would be seen by an observer looking along the vector of the light source and can be used as an intensity value for the element when it is displayed in 2D. The scaling of this value, necessary to create integer values of pixel intensities, is a form of representation of the texture and hence reflectivity of the object.

With reference to figure 4.4 the points A, B and C can be expressed by two vectors, $\mathbf{a}$ and $\mathbf{b}$ where

$$\mathbf{a} = \overrightarrow{CA},$$

and

$$\mathbf{b} = \overrightarrow{CB}.$$

Firstly, the normal vector to this plane is required. This is given by

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a}|\,|\mathbf{b}| \sin\theta.\hat{\mathbf{n}},$$

where $\hat{\mathbf{n}}$ is the unit vector normal to the surface and θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. This can then be reduced to get

$$\mathbf{a} \times \mathbf{b} = (a_y b_z - a_z b_y)\mathbf{i} + (a_z b_x - a_x b_y)\mathbf{j} + (a_x b_y - a_y b_x)\mathbf{k}. \tag{4.1}$$

By considering the vector $\mathbf{l}$, which is the vector from the light source, L, to the surface element, it is possible to find the angle, ϕ, between this vector and the normal vector found in equation 4.1 using

$$\mathbf{l} \cdot (\mathbf{a} \times \mathbf{b}) = |\mathbf{l}\,|\,|\mathbf{a} \times \mathbf{b}| \cos\phi$$

and hence

$$\begin{aligned} \cos\theta &= \frac{\mathbf{l} \cdot (\mathbf{a} \times \mathbf{b})}{|\mathbf{l}\,|\,|\mathbf{a} \times \mathbf{b}|}, \\ &= \frac{a_x d_x + a_y d_y + a_z d_z}{\sqrt{a_x^2 + a_y^2 + a_z^2}\sqrt{d_x^2 + d_y^2 + d_z^2}}. \end{aligned}$$

Where $\mathbf{d}$ is the vector product $\mathbf{a} \times \mathbf{b}$.

Quite clearly this is very computationally intensive. Even for powerful computer facilities, full frame generation can take as long as 30 minutes.

4.1.4 Hidden line and hidden area removal

There are three ways traditionally used to remove hidden parts of 3D objects: The hidden sections can be calculated from the coordinates of the surrounding objects; some visual trick can be used that makes use of the two dimensionality of the screen or the resolution of the screen can be employed to achieve the required effect.

The first of these involves large amounts of computation even for line drawings. Every line must be intersected with every other line that could be in front of it and, if they are found to be overlapping in the particular projection being used, then only the visible parts are displayed. The extension of this to area intersections is used to allow shaded hidden area removal. Some reduction in the computational requirements of this approach have been achieved but it is mostly the application of surface

drawing techniques to more complex 3D rendering that has drawn the attention of researchers [74, 75].

Both of these methods are very versatile, and can produce correct images from any viewing point and any orientation. However, even for very powerful computer facilities this sort of calculation takes a long time and only supercomputers can achieve speeds approaching real time rendering, and then only for simple images. Techniques such as sorting the image data into depth related order can help to limit the number of calculations needed but such algorithms are still too complex for effective hardware implementation.

Because of this complexity a number of simplification techniques have been developed. If it is possible to draw the areas that make up a shaded surface in such a way that once drawn an area can only be obscured, and can never itself obscure areas that are drawn later, then it is possible to draw the image starting with points furthest away and then moving to those closer in to the observer. Each area will now be drawn over any sections that are behind them, thus obscuring them. The area may in turn be obscured by something that is drawn later but cannot appear in front of such a later addition. Once all the shading has been achieved, the correct 'solid' image is displayed.

This same technique cannot be applied to wire frame drawings because there are no shaded areas to blank out hidden parts. One solution is to blank out the areas between the lines by pretending that they are shaded in black. This works well, but still requires very much more calculation than the wire frame drawing on its own would have needed. However, if the surface is drawn from front to back as with the shaded versions the following algorithm will work just as well and has the added advantage of requiring very little computer overhead.

This algorithm will display wire frame surfaces provided the projection conforms to the above restrictions.

```
1. Draw the line that is furthest from the observer.
2. Draw any line that is in the same plane as this one.
3. Start to draw the next furthest line.
4. Each point that is to be drawn has a position on the
   screen given by X=x,Y=y, compare this point's y
   coordinate to the y coordinates of any other points
   plotted on the same vertical line (i.e. on the line X=x).
5. Plot this point only if there are no points above it or
```

```
   there are no points below it.
6. Repeat this for all lines in the same plane as this one
7. Repeat this for all remaining planes in the image.
```

As can be seen the image is built up from back to front making use of the resolution of the display to simplify the calculation.

This method is the basis used for the hardware implementation that will be covered in the following sections. All the data used comes from a grid such as the cells on the surface of a CCD. If the image is displayed such that the x-axis of the CCD is parallel to the horizontal axis of the screen, and the intensity signal for each point is parallel to the vertical axis, then there is one more simplification available. If the image is built up by starting with the lines connecting $(0, 0, z_0)$ and $(1, 0, z_1)$ then the only lines that are in the same plane as this one are those with y coordinate of 0. It is then possible to build up the image as a set of lines from

$$(0, 0, z_0) \rightarrow (1, 0, z_1) \rightarrow ... \rightarrow (n, 0, z_n)$$

and to know that all of these lines are in the same plane and hence do not overlap.

This technique is not true for lines drawn in the planes where x is a constant, such as the lines from $(0, 0, z_0)$ to $(0, 1, z_1)$, as these are not drawn in the depth first order. Because of this it is necessary to use only the lines parallel to the x-axis and ignore those parallel to the y-axis. As will be seen in later examples this does not detract significantly from the appearance of the image.

4.2 Fast Software Solutions

4.2.1 Simplification of viewing angle

When the mapping from 3D to 2D is done, it is usually necessary to define both a viewing position of the eye and a position of the screen in the coordinate system of the surface to be displayed. The screen coordinates are deduced by finding the point at which a line connecting the viewing coordinates to the surface element intersects with the screen, as described earlier. For drawing purposes however this mapping can be viewed as simply the incorporation of the real world y axis into the x and z axes to form the x and y axes of the screen. This is the way a normal technical drawing is done.

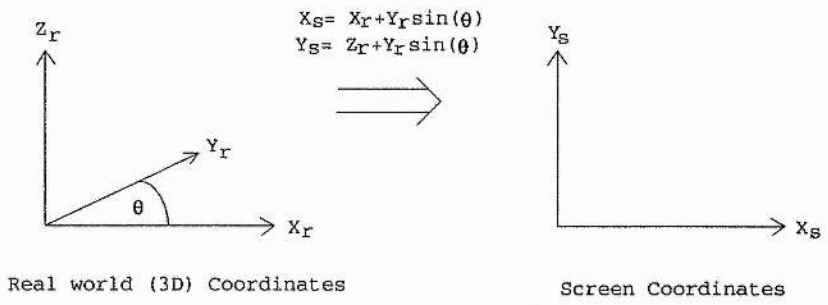


Figure 4.5: The screen and world coordinates

Figure 4.5 shows the representation of the x, y and z axes in this form. The standard projection is obtained when the angle θ (as measured on the page) is 30°. However, in general the mapping follows the form

$$\begin{aligned} x_s &= x_r + y_r \sin(\theta), \\ y_s &= z_r + y_r \cos(\theta). \end{aligned}$$

Where (x_s,y_s) and (x_r,y_r,z_r) are coordinates in the screen and real world (3D) coordinates respectively.

This poses the first obstacle to finding a hardware solution, namely the need for trigonometric functions, which are inherently difficult to implement in hardware. For practical purposes the projection can be fixed, and the sine and cosine values can be hard-coded as two real numbers. This still presents complex hardware problems. However, if the projection is changed so that $\theta = 45°$ then the y axis of the real world coordinates will be displayed on the screen as a diagonal line defined by the equation $y_s = x_s$. The 45° projection defines $\sin 45° = \cos 45° = 0{\cdot}707$. By approximating this value to 0·75 it is now possible to calculate the mapping by a simple multiplication by $\frac{3}{4}$. Hence, with a 6% error it is now possible to define the mapping as

$$x_s = x_r + \frac{3}{4} y_r,$$

$$y_s = z_r + \frac{3}{4} y_r.$$

Since the minimum change that can be made in y_s is one pixel, the approximation of $\sin 45° = 0{\cdot}75$ cannot be used for y values that are not multiples of 4. this means

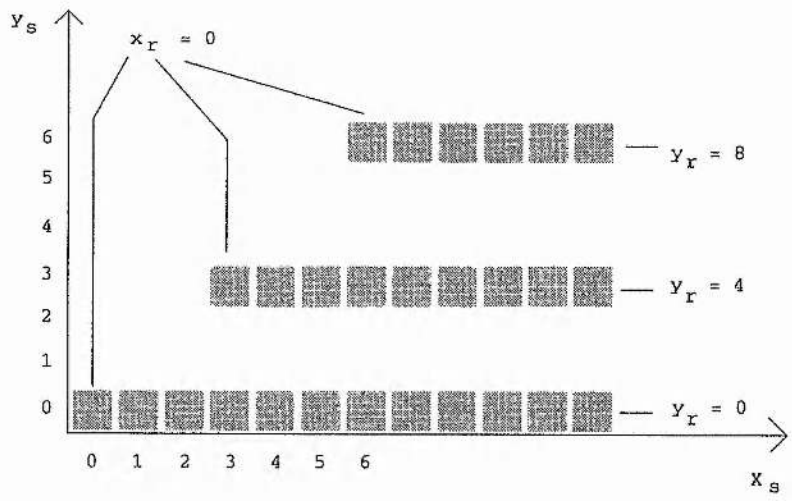


Figure 4.6: Part of a surface representation with constant z values.

that some information must be lost if a VGA type display is to be used as the 640 by 480 resolution prevents the required scaling in either axis that would allow all the available data to be used. Because of the simplification of mapping x_r pixels to x_s pixels on a one to one basis, it was decided to simply use every fourth row from the CCD image.

In this way $(0, 0, z_0)$ maps to the screen coordinates $(0, z_0)$ and the next pixel in the y_r axis that can be drawn is $(0, 4, z_4)$ which maps to $(3, z_4 + 3)$. For a plane defined by $z = 0$, this will appear as in figure 4.6.

Visually it was found that the multiple of three for the y_s axis was not essential. As can be seen from one of the later examples (e.g. figure 4.13) the use of an offset of just one pixel gave very good results. The larger the gap the more 'above' the surface the observer appears to be. Because of this new simplification, the surface can be plotted by taking every fourth horizontal line from the CCD image and displaying it offset by one pixel in both the x_s and y_s axes for each line plotted.

Because the final aim is to produce a camera system that outputs general surfaces this fixed projection would appear to be a serious limitation. However, in order to change the view it is simply necessary to rotate the camera about its longitudinal axis rather than change the image projection. This will have the effect of rotating the surface without changing the projection method.

The need to get images from a camera did however cause one problem. The lines

parallel to the y axis can only be drawn when the next line data is available. Examples were examined to determine alternative methods and these showed that given sufficient numbers of horizontal lines, the vertical lines could be ignored and the images could still be correctly interpreted. The real life example given later illustrates this quite well.

The initial complex surface generation has now been simplified to looking at every fourth y_r value and hence the mappings

$$x_s = x_r + \frac{y_r}{4}, \qquad (y_r = 0, 4, 8)$$

and

$$y_s = z_r + \frac{y_r}{4}.$$

As can be seen a hardware counter can provide the values of y_r and x_r, and as z_r values are read out from the CCD the resultant x_s and y_s can be found by using a simple adder circuit to combine the y_r value with the x_r and z_r (*intensity*) values respectively.

4.2.2 Shading algorithms

Simplification of the shading algorithms is more complex. Two techniques were tried, using visual properties of surfaces to allow simplifications in the calculations.

The first method used was to display the shade as a function of the area taken up on the screen. This is based around the fact that when these areas are large, the gradient of the surface is also large which in turn means that the angle to the observer is smaller. Unfortunately, this only applies when the areas being considered are angled in only one plane. With a surface there are two axes of rotation and hence it is possible to invalidate this algorithm.

It was found that when implemented on a 12MHz 80286 PC, a 50 by 50 grid took 42 seconds to display using a grey-scale image of only 16 shades. An example of this type of shading is given in figure 4.7.

Figure 4.8 shows the typical situation. The shade of the quadrilateral is improved by splitting it into two triangles and dealing with each separately. The area of an arbitrary triangle is given by defining

$$a = \sqrt{(A_x - B_x)^2 + (A_y - B_y)^2} = \left|\overrightarrow{AB}\right|,$$

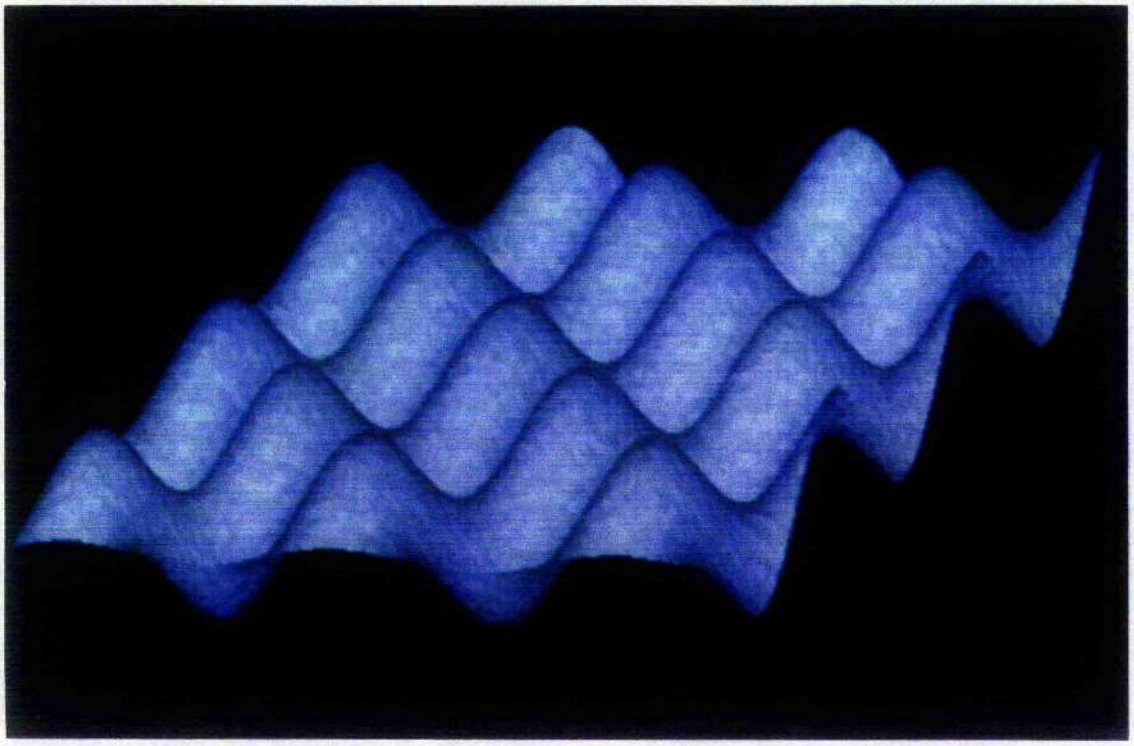


Figure 4.7: Results of using area to define shade.

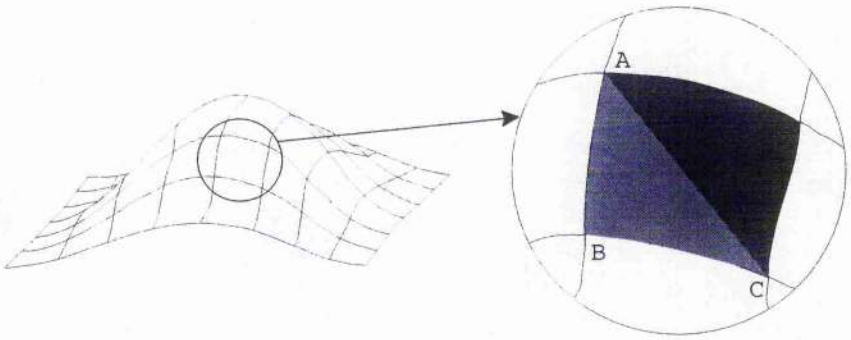


Figure 4.8: Surface generation using shade proportional to area.

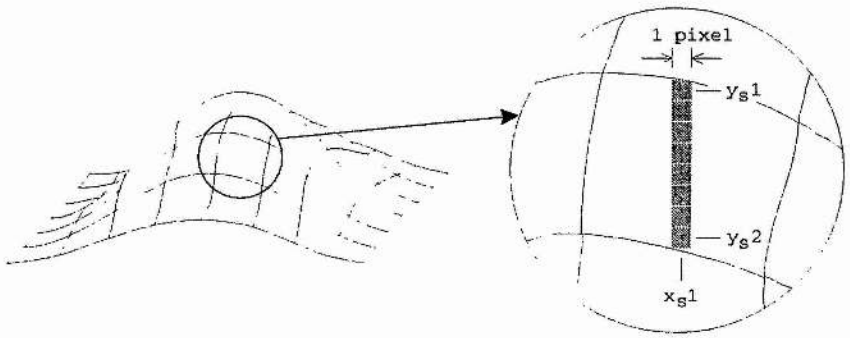


Figure 4.9: Area calculation by using single pixel widths.

$$b = \sqrt{(B_x - C_x)^2 + (B_y - C_y)^2} = \left|\overrightarrow{BC}\right|,$$

$$c = \sqrt{(C_x - A_x)^2 + (C_y - A_y)^2} = \left|\overrightarrow{AC}\right|,$$

$$s = \frac{a+b+c}{2}$$

and

$$\text{area} = \sqrt{s(s-a)(s-b)(s-c)}.$$

Unfortunately this is still very hard to calculate in hardware. One solution is to make use of earlier simplifications in the wire frame drawing. Here each horizontal cell of the CCD is mapped to one pixel on the screen. This means that the width of the quadrilateral is only one pixel. Hence the area becomes the difference between the y_s value of this pixel and the corresponding y_s value of the pixel in the previous row that is positioned at the same x_s coordinate. This is shown in figure 4.9 where the area is proportional to $y_s1 - y_s2$. When implemented on the PC this algorithm reduced the rendering time to 30 seconds for a full CCD image rendered on the display hardware using 256 grey shades. The results in figure 4.10 are from an image that was generated by reading the frame buffer into the PC and then sending the surface image to the graphics display hardware after processing it for shading.

This example also illustrates one other change. It was found that if the shading was gradually reduced from front to back, then an improved result was obtained giving the impression of reduced intensity with distance from the screen.

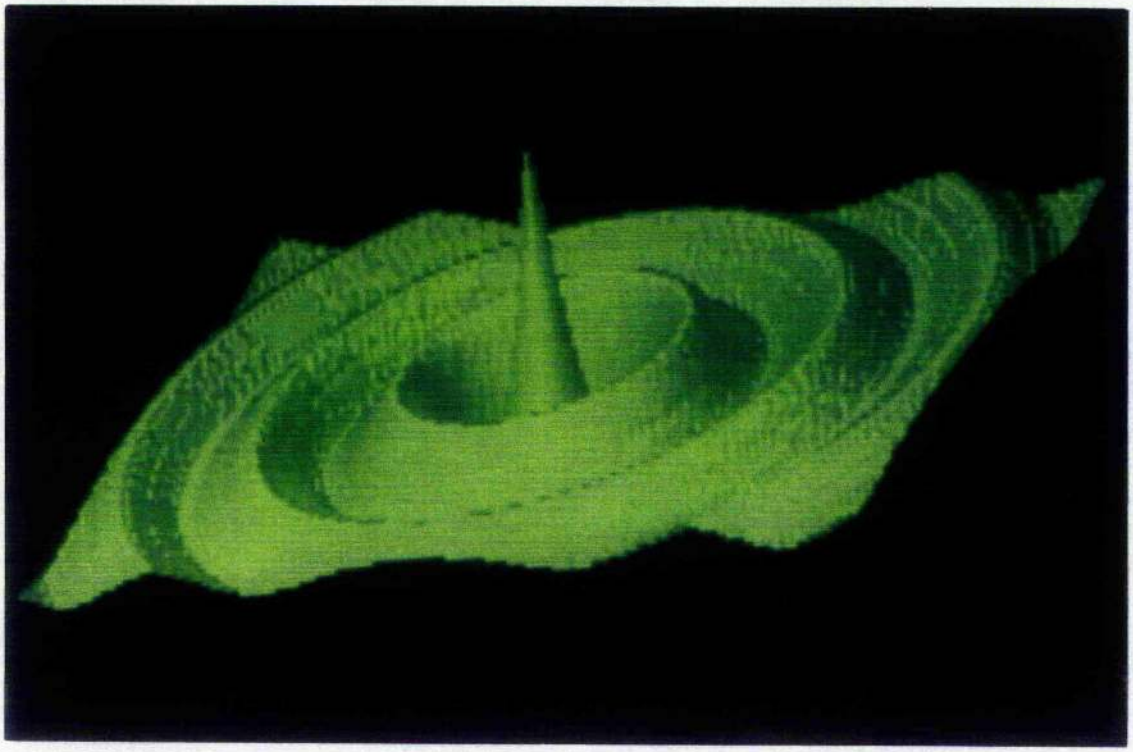


Figure 4.10: Results of single pixel width area calculation.

Visually, these techniques gave very good results for most simple surfaces. The results can be compared to figures 4.11 and 4.12 which were generated by the vertical difference and the full cosine reflection methods respectively.

4.2.3 Hidden line removal

The simplified projection described earlier makes it possible to modify the hidden line removal algorithm to make it simpler to implement in hardware. The projection requires that every new horizontal line taken from the CCD image has a start point one pixel to the right of the previous lines start point. If it is decided to map one CCD cell to one pixel on the screen then it is no longer necessary to connect these two pixels with a line. In most cases they will be very close together on the screen and only when high gradients are being displayed will they be far enough apart for the missing pixels to become evident. The results of this can be seen in figure 4.13 where only the peak areas in the centre show significant reduction of quality. It should be noted that this effect is very much less pronounced in the hardware versions as the oscilloscope beam is continuous and hence 'fills in' the missing lines.

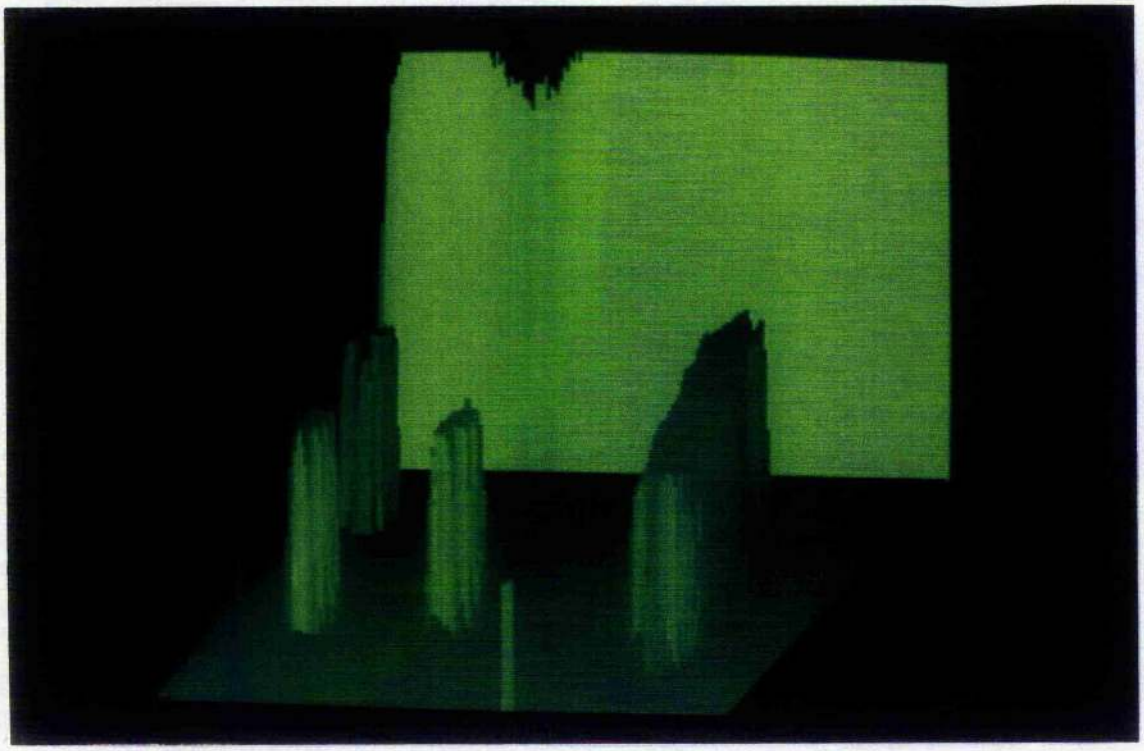


Figure 4.11: Brass cutout image shaded using vertical differences.

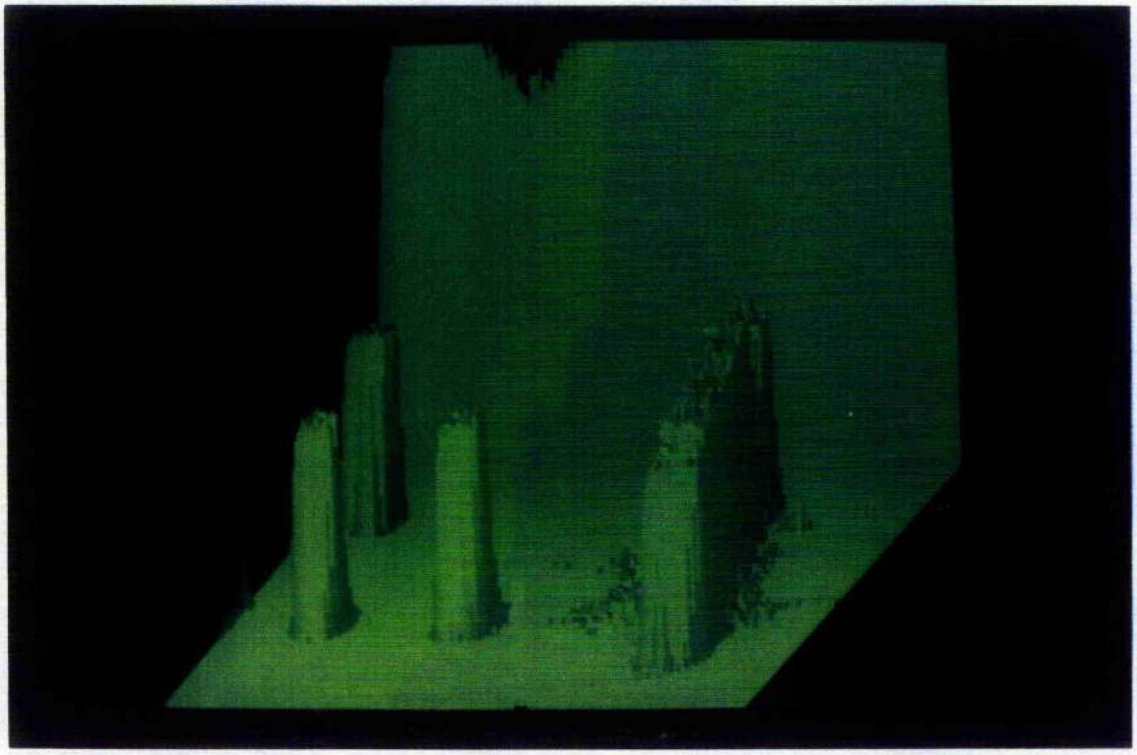


Figure 4.12: Using light vectors to generate a full surface representation.

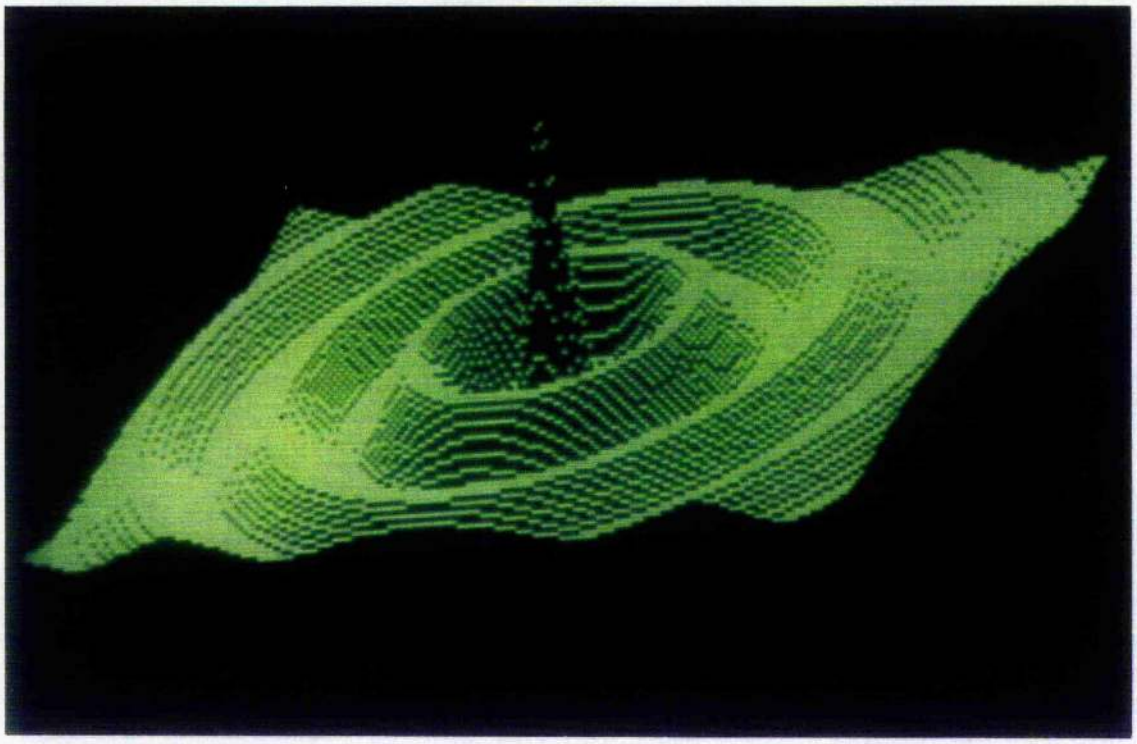


Figure 4.13: Example wire frame using only the x axis pixels.

There are now only hidden pixels rather than hidden lines to be removed. The removal of these pixels is done by using x_s as an index into an array where the greatest value of y_s used at this value of x_s is stored. If a pixel has a y_s value greater than the stored value then it is displayed and the stored value updated. If the pixel has a y_s value that is less than that stored then it is not displayed. This only takes care of the top half of the image. It is possible, as the previous algorithm illustrates, to see the under-side of the surface. This is achieved, in a similar way, by using another array that contains the minimum values of y_s. Now, for any given x_s, y_s pixels are displayed when they are less than the stored value.

When executed on a PC reading the CCD frame buffer, the rendering time was reduced to 2·8 seconds for this method.

4.3 Basic Hardware Solutions

4.3.1 Adaptation of algorithms to hardware

In order to design a real time surface generator certain parameters had to be set. Firstly a refresh rate had to be established. Normal PAL TV uses 50Hz but tends to

generate eye strain and can produce visible flicker in some cases. Because of this it was decided that the screen refresh rate would have to be closer to 100Hz to give a good quality display.

Fortunately the image update frequency does not need to be this high. The human eye tends to be only able to spot changes over about 100ms or so, hence with a stable image provided by the 100Hz refresh rate, the image update rate can be as slow as 10Hz.

The type of display also had to be defined. The normal, electromagnetic, displays are inflexible and very limited in their capability so it was decided to make use of the higher bandwidth and beam gating capability of an oscilloscope's electrostatic display. Because of the variability of the vertical input gain, it was also possible to have a scaling of the image vertically. A larger input gain leads to a larger vertical offset between rows, thus the correction for the 4 line vertical spacing needed to meet the theory could be set quite accurately.

The display selected was an 80MHz Hameg oscilloscope. The flyback can be accurately controlled on this device and the availability of trigger delay inputs meant that the unit could be better tailored to provide special display requirements.

In order to update the image being displayed, without causing flicker or loss of picture, it was decided to build the generator with a double memory system. The data in one memory block being output to the screen while the other is being set up. A single write instruction from the Z80 processor will then flip the memory banks over and the new data will be displayed and the old will become available for updating.

Figure 4.14 shows the block diagram of the basic hardware. As can be seen the ability to turn off the display when no pixels are to appear was achieved by using the most significant bit of the data store as an output driving the beam gate input of the oscilloscope. This made it possible for the oscilloscope beam to be turned off at any point in the output waveform, at the expense of only allowing 7 bit data to be displayed.

The memory addresses were initially accessed by a counter that counted sequentially through the RAM addresses 0 to 32k. This was set up as 512 pixels per row and 64 rows so as to enable a one to one mapping from CCD cells to output pixels. There are 388 pixels on a CCD row and this choice of RAM configuration gives a maximum y value

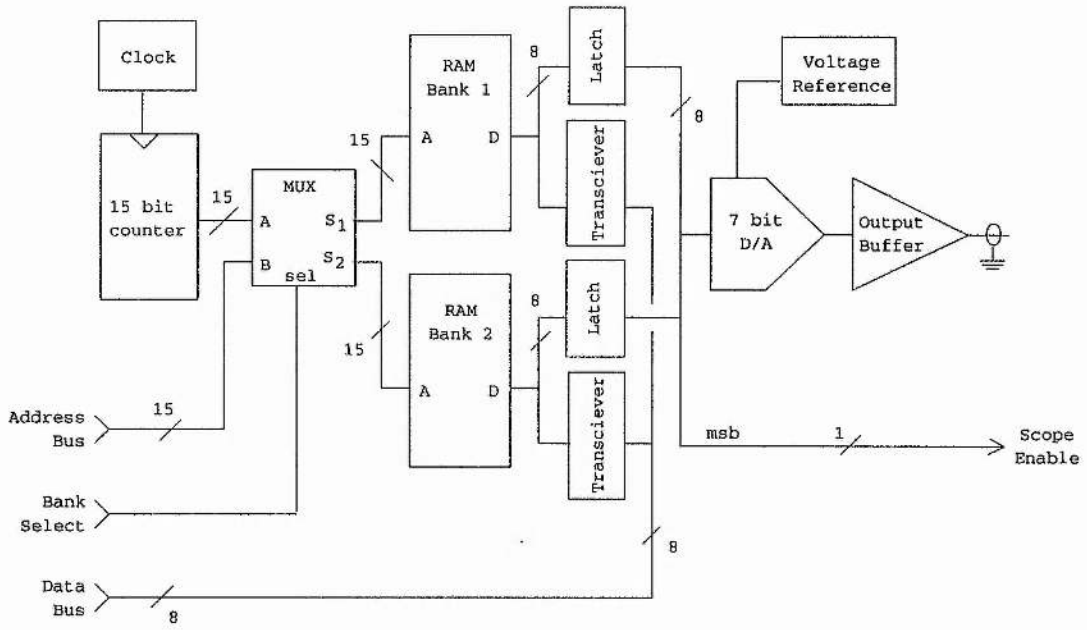


Figure 4.14: Block diagram of the initial hardware.

of 64. This maps to a maximum x_s value of $388 + 64 = 452$, hence a 9 bit address is required.

As there are 288 lines on a CCD, and every 4 lines are to be displayed, the ideal would be to have 72 lines displayed. Unfortunately the 32k limit only allowed 64 lines but, as can be seen from the end results in figure 4.19 this did not significantly degrade the image.

The circuit was developed to attach to the frame grabber, camera and display hardware developed in chapter 1. This allowed the board to be directly accessed by the Z80 processor and indirectly by the PC controller. The RAM was accessed directly via the Z80 address and data buses, and the two banks of RAM were switched via status line outputs from the processor board, thus allowing the switch from one bank to the other to be done in a single write operation as defined earlier. This board also provided all the timing signals, such as vertical blanks, that are needed for such a system. Were this hardware to be provided as an in line unit for a normal video signal it would be necessary to extract the vertical and horizontal synchronization signals and use these to detect every fourth row and put it into the surface generating RAM.

Another advantage in using the frame grabber hardware was that the RAM layout

used for the frame storage was four way interleaved and hence transfer of every fourth line for surface rendering could be done from only one physical RAM chip.

4.3.2 Multiplexors

The normal TTL multiplexor, the 74LS257, only provides one 4-bit output set, the selection input routing either the A or the B inputs to the outputs. For the purposes of the bank selection a multiplexing system was devised that would allow a control input to either route A to S_1 and B to S_2 or to route A to S_2 and B to S_1. This was engineered by a 16v8 GAL, with logic outputs S_1 and S_2 defined by

$$S_1 = (\text{ctrl} \wedge A) \vee \left(\overline{\text{ctrl}} \wedge B\right),$$

$$S_2 = \left(\overline{\text{ctrl}} \wedge A\right) \vee (\text{ctrl} \wedge B).$$

With the 20 pin package it was possible to have four such sets all driven from a common control input, replacing the two 16 pin devices needed to do this in discrete TTL.

4.3.3 Output filters

The output of this memory was now sent to an D/A unit, the PNA7518 [5], and then through a buffer amplifier and on to the oscilloscope voltage deflection input. Very quickly it became evident that the bandwidth of the scope was too good for our purposes as the beam stepped up to each value too quickly and the expected filling in of the lines between each pixel (section 4.2.3) did not happen. This problem was overcome by filtering the display output. An active, 2 pole, Chebyshev filter was used for this purpose.

With reference to figure 4.15, the time constant is given by

$$rc = \frac{1}{2\pi f_n f_c}, \tag{4.2}$$

where f_0 has the dimensions of s^{-1}, f_n is dimensionless,

$$\left.\begin{array}{rcl} f_n & = & 1{\cdot}231 \\ k & = & 1{\cdot}842 \end{array}\right\} \text{ for 0·5dB ripple,}$$

$$\left.\begin{array}{rcl} f_n & = & 0{\cdot}907 \\ k & = & 2{\cdot}114 \end{array}\right\} \text{ for 2·0dB and higher speed tailoff}$$

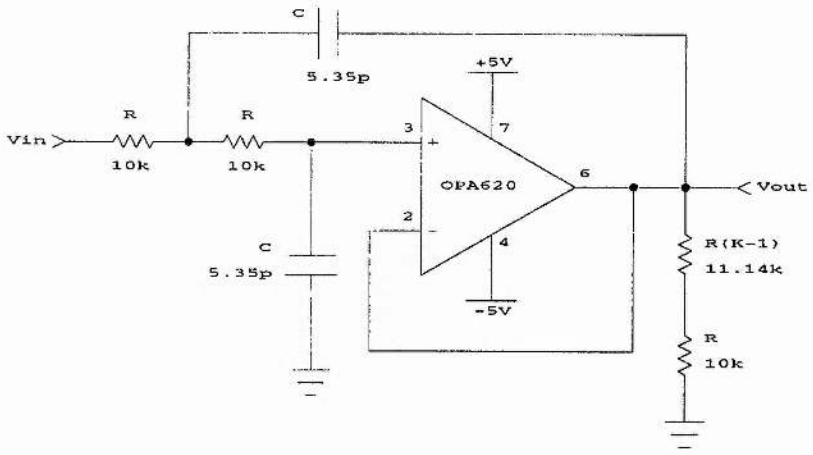


Figure 4.15: The Chebyshev filter

and k is the voltage gain of the amplifier stage.

As the circuit is using 1024×32 bytes per frame and 100 frames per second the clocking speed for the pixels is 3·28MHz. For the 2·0dB tail off, equation 4.2 gives $rc = 5{\cdot}35{\times}10^{-8}$s and, choosing r to be comparable to the op-amp input impedance, say 10kΩ, c is now found to be 5·35pF and $r(k-1) = 11{\cdot}14\text{k}\Omega$.

Along with a simple, low noise, voltage reference for the D/A this completed the analog part of the circuit. Because of the video rate operation of this circuitry, the voltage source had to be very stable. An MP5010 1·22V fixed voltage reference was used. This fed an OP-27 low noise instrumentation amplifier, wired as a non-inverting gain stage. The complete circuit diagram for the first stage surface hardware is given in inserted figure VII.

4.3.4 The firmware interface

Before an image was generated the processor waited for a vertical blank period and then did a frame grab. Once available, this data was used to fill the surface RAM. It was then necessary for the processor to do all the mappings from real world to screen coordinates.

Each line seen on the display represented 512 pixels. For the first line of the image

the processor had to fill in the first 388 bytes with the corresponding first row of the CCD frame store. As the most significant bit was needed for the gate output, this data had to be shifted down by one bit (i.e. divided by 2). It also had to be shifted by one bit again so that the vertical component of y_r could be added in to give the y_s value. For the rest of the 512 bytes in this row the processor put in a value with the most significant bit set (in this case 128). The hardware monitored this bit and, when set, the beam was turned off. In this way the line being drawn would appear to end part way across the screen.

For the next 512 bytes, the processor set the most significant bit of the first byte, thus achieving the addition of the y_r component in the horizontal axis. The processor then moved the fourth line of image data into bytes 2 to 389 of the display, performing the two shifts and adding in the $\frac{3}{4}y_r$ value (i.e. for this line adding in the value 1) to each byte as it placed it in the memory. Finally the processor set the highest bit of all the bytes in the rest of the row to prevent them from being drawn.

This process was repeated for each line until the entire 64 lines of the frame buffer had been sent to the display. Once finished the processor changed over the memory banks and the new data became visible.

Finally the whole procedure was then repeated for the second memory bank and then back to the first, thus providing continuous update of the image.

4.3.5 Hidden line calculations in software

The simplest method of achieving the hidden line removal was to read in the frame buffer data into the PC, carry out the hidden line routines and then output the data to the surface generator via the Z80. The routine to do this achieved good results but was slow as all the transfers between the PC and the Z80 are very time consuming. The best results achieved frame rendering in 40 seconds. This was done by using only integer calculations and optimizing the code that read in the CCD data so that it only accesses every fourth line.

4.3.6 Results from a Z80 implementation

Once the code had been developed in PASCAL, it was now possible to rewrite it in Z80 assembler thus considerably speeding up the process. Running on a 6MHz processor,

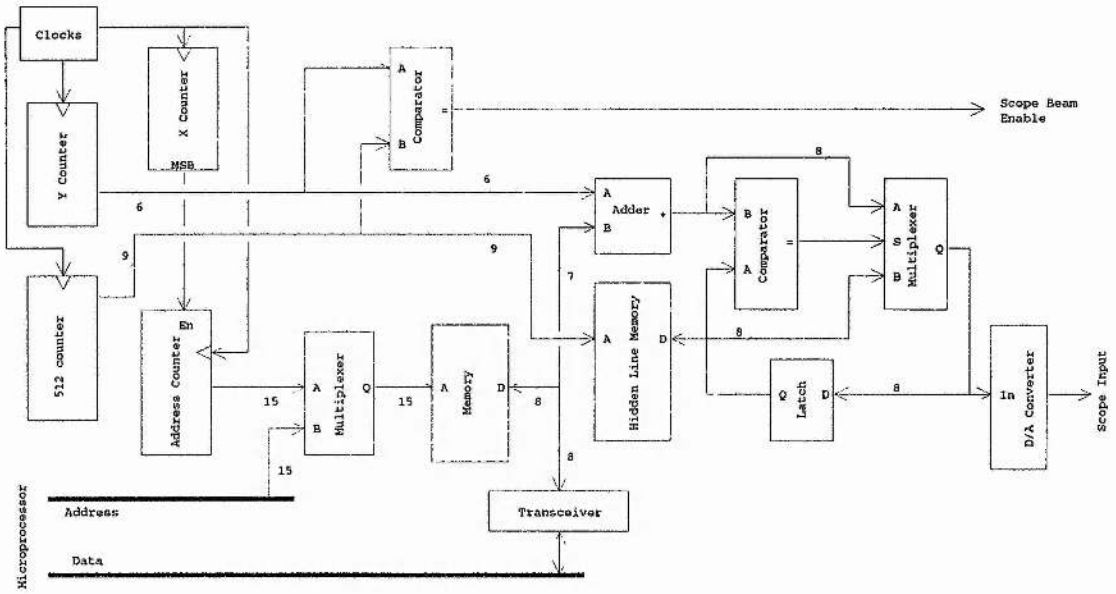


Figure 4.16: Block diagram of the full surface generator.

the rendering, timed over 100 frames, was now reduced to 1·066 ± 0·005 seconds per frame when the Z80 was running in polling mode, and when dedicated to frame rendering this time was reduced to 0·954 ± 0·005 seconds per frame.

4.4 Hidden Line Hardware

4.4.1 Adaptation of hidden line algorithms

Having developed the basic display hardware the next stage was to look at implementing the hidden line algorithms in hardware. The software described earlier consists of three parts: the addition of the y_r component into both x_r and z_r (to generate x_s and y_s) and the hidden line removal. These also convert into three functional blocks of hardware as illustrated in figure 4.16

In order to add in the y_r component into x_s two counters and a comparator had to be added to the original circuit. One counter counts up from 0 to 511. This represents the pixel count for a single line. This counter then triggers the second counter to increment the number of lines displayed. This is similar to the original system, however, a comparator is used to compare the output of the pixel counter to the output of the line counter. By making the circuit inactive while the pixel counter

is less than the line counter, a time delay which is proportional to the line number, is produced. This provides the left hand offset of the image line, effectively doing the mapping of y_r into x_s.

When the comparators $\overline{\text{a=b}}$ output goes active, it is latched to prevent further changes (which would occur every 64 clocks as only 6 bits are used for the comparison) and then fed into an AND gate along with the clock signal. This permits the original pixel clock signal to get through to two more counters. One of these counters provides addressing for the data RAM and will output successive values to the display. This is only reset at the end of a full display output. The other counter is reset before each line is output. It counts the input pulses until it reaches 388. When this value is reached, the counter is reset and a signal inhibits the clock pulses going into the address counter. The pixel counter will continue counting until it reaches 511, thus putting in the time delay needed to make each output line the same duration.

The ability to stop the pulses getting through to the address counter means that it can simply be left where it is and will start from this new point when the next 388 pixels are to be read out. This allows the data from the frame buffer to be transferred all in one block as the break up of lines for this circuit is now the same as it is for the frame buffer (i.e. both are simply contiguous). Unfortunately this does not give more y_s output capability as the next possible increment to the number of lines would be 128 and there are only 72 available from the CCD. The large delay that this would result in would add flicker to the display as it would also halve the refresh rate.

One problem that can now be solved is the presence of the saturation line that appears when the CCD image is shifted down. For the normal camera image this appeared as a bright line at the top of the screen, but, if this were to appear in the surface representation then a 'wall' would appear in front of the required image. As the frame buffer is laid out in an interleaved system the image does not have to be taken from the lines 0, 4, 8 etc. and can in fact be taken from the image at lines 3, 7, 11 etc. by taking the image from the fourth interleaved memory block, thus ignoring the saturated lines in the image.

When the final, 64th line has been output a single blank line is sent to the display. This allows the hidden line circuitry to be reset by writing zeroes into the memory. After this everything is reset to begin the output cycle again.

The generation of y_s is slightly harder as the correct fraction of y_r has to be added to the intensity (z_r) value for every pixel displayed. One possible way to do this would be via an analog summing amplifier driven from the output of the z_r D/A and also from the output of a D/A attached to the y_r counter described earlier. This system, while effective, is very prone to noise and also makes it hard for the design of a hidden line circuit which is an inherently digital process.

Instead of the analog system a more complex digital summer was used based upon two 74LS283 4 bit adders. These adders were set up to add one six bit to one seven bit value. The six bit input being the y_r counter (0 to 63) and the seven bits being supplied by the most significant seven bits of the intensity data for this pixel (z_r). In this way the adder output is guaranteed to be within the range of its eight bit output. Because the top seven bits of the image data have been used, the transfer of data from the frame buffer to the surface RAM is now a one-to-one transfer with no data processing being done, the hardware simply ignores the least significant bit of data.

The output of this summing circuit, when fed via the D/A and op-amps, gave the full surface with no hidden line removal. This provided a good test point for the circuit and also made it possible to add a switch that would either enable or disable hidden line removal by feeding either this output or the hidden line output into the D/A and buffer stages. Figure 4.17 shows an example output of this stage.

The last requirement was to remove the hidden lines. This was based very closely on the program version described earlier, and consists of a memory that acts as the test array in the code, a comparator that selects either the stored value or the new pixel value depending on which is the greater, and a set of latches to select which of these should be output to the oscilloscope. The output of the comparator also provided the gate enable signal for the oscilloscope. This output was latched along with the data so as to disable the beam if the y_s value it is receiving is a hidden value.

This hidden line circuit operates by feeding the value of the 0-511 pixel counter into the address inputs of a RAM. Initially set to zero at the start of a frame, this RAM stores the highest y_s value used for this counter output (i.e. for this value of x_s). The data outputs of this RAM are then sent to the comparators with the output of the summing circuitry described earlier. If the stored value is greater than the summed output then this, along with the comparator signal, is latched into the output register

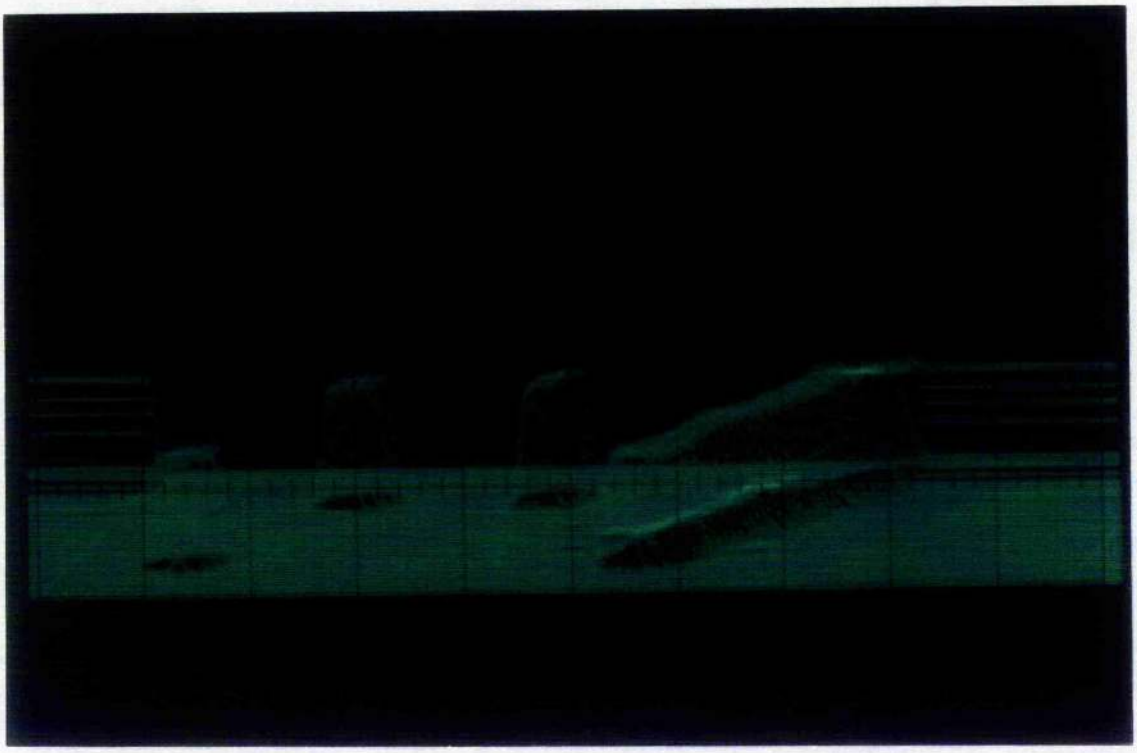


Figure 4.17: Surface rendering without hidden line removal.

on the negative edge of the x_s counter. If the stored output is less than this summed output then it is the summed output that is latched.

During the negative half cycle of the x_s clock the value of this latch is written back into the RAM. Either the new maximum y_s is stored or the old value is written back in. During the memory erase period this latch is disabled and the data bus is forced to zero by the pull down resistors. Hence zeroes are written into the memory on every write cycle for the entire 512 pixels.

On the next rising edge of the clock the contents of the comparator latch are transferred into another latch to generate a stable input to the D/A converter and the analog circuits.

Because of the use of a full 7 bits of data from the image, it was no longer possible to use the RAM as a means of turning the display off when no pixels were to be displayed. The comparator output provided an off signal for use when there were hidden parts of the display but did not solve the problem of what to do about the edges of the display where there was no surface to be shown.

Figure 4.18 shows the resultant surface when no means of turning the beam off was

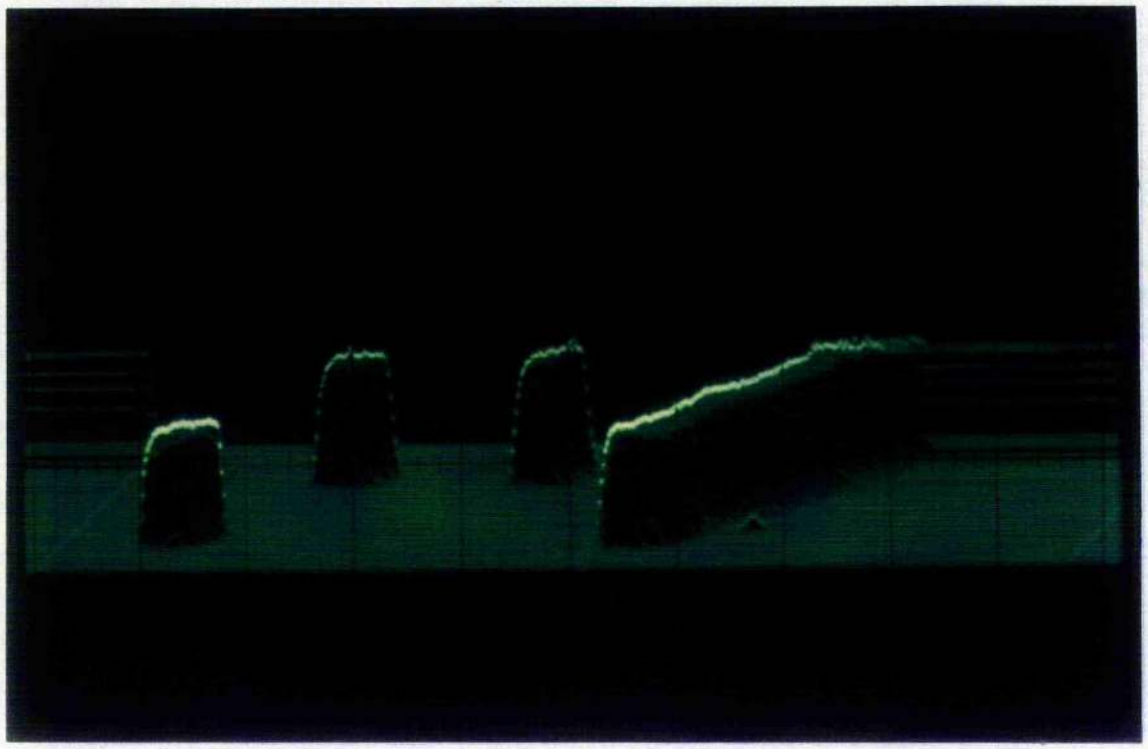


Figure 4.18: Example output without using any beam gating.

used. This shows the brightening of the image where the beam repeated parts of the image where hidden line effects were detected. It also shows the 'mess' that is produced outside of the image due to there being no means of turning the beam off when no image should be displayed.

The presence of the address counter enable line, available from the pixel comparator stage, provided a suitable beam gate to remove this unwanted part of the image. Figure 4.19 shows the effect of putting this signal, along with the comparator output, into an OR gate and then into the beam gate input of the oscilloscope. The removal of the bright spots and the non-surface parts of the image gives a much more pleasing picture.

Most of the logic gates for this circuit were implemented within a 16v8 GAL. Figure 4.20 shows the logic equivalent of this device. Because of the lack of pull up resistors in GAL outputs, the GAL could not be used to generate a reliable mono-stable, this was implemented in discrete logic. Inserted figure VIII shows the final circuit diagram of the surface generator.

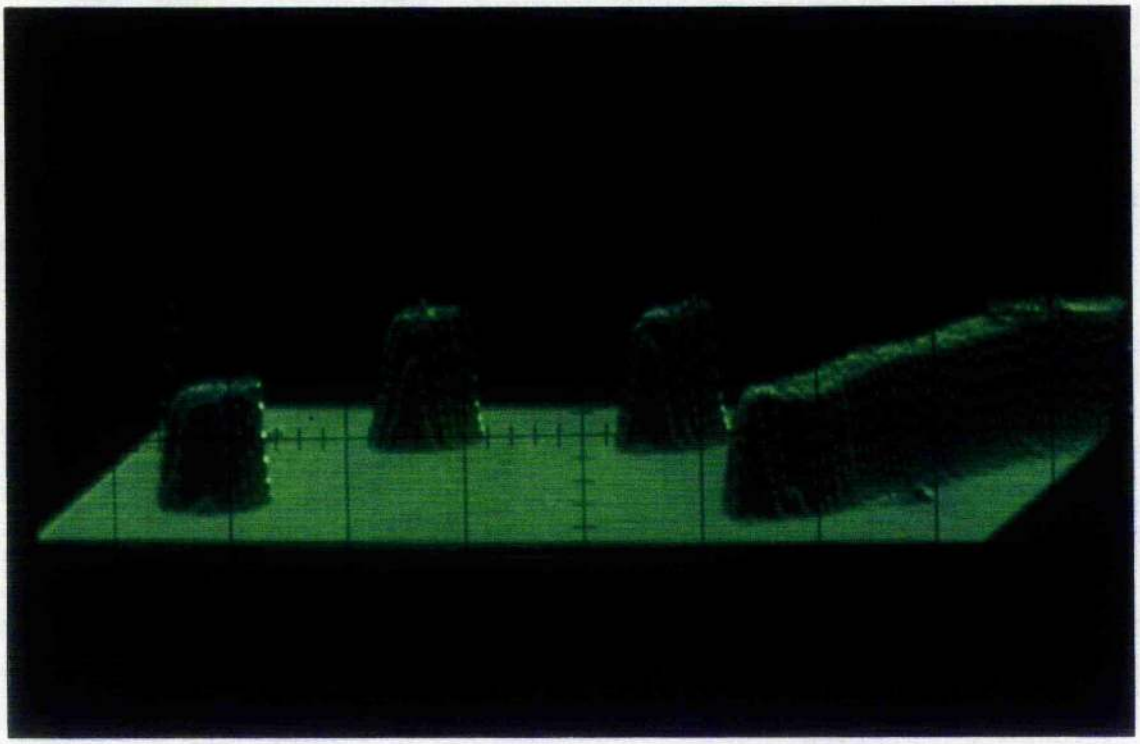


Figure 4.19: The effect of gating the beam.

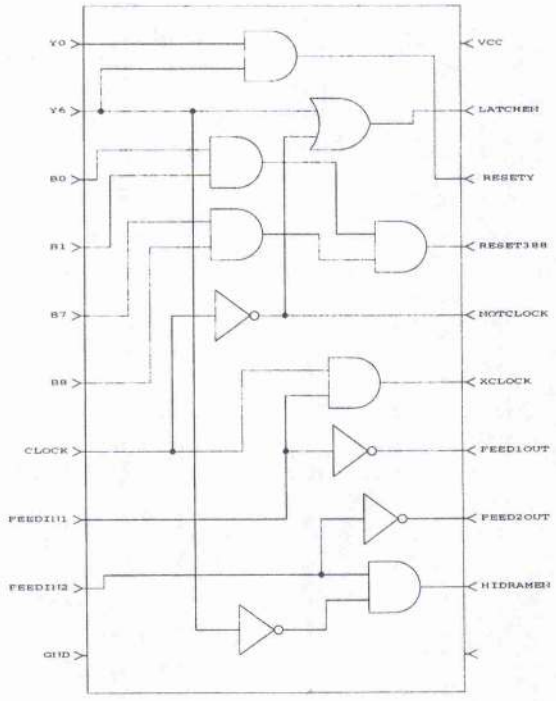


Figure 4.20: The equivalent logic of the 16v8 GAL.

4.4.2 Results

Figure 4.19 shows the output created from the test image used to evaluate the camera system developed in chapter 1. It illustrates a number of points. The low background noise level can be seen in the flat parts of the image, and the dynamic range of the output is clearly illustrated in the high gradient areas of the image. This picture was taken while the unit was running in real time. Slight movements of objects within the image could be clearly seen on the display and the resulting images were very stable. One point to note here is the enhanced detail that surface representation of this type provides. The rough edges of the cutout are noticeable and the extent of the light spread out can be clearly seen.

4.4.3 A DMA solution

Because of the need for high speed transfers, the Z80 system used in the previous circuits was found to be too slow. Using the block load, LDDR, command to move memory blocks, the transfer rate is given by

$$\begin{aligned} \text{rate} &= \frac{1}{\text{frame time}} \\ &= \frac{\text{CPU Clock Frequency}}{\text{No. Bytes} \times \text{T states per LDDR}} \\ &= \frac{6 \times 10^6}{388 \times 64 \times 21} \\ &= 11{\cdot}51 \text{ frames } s^{-1}. \end{aligned} \tag{4.3}$$

Though good, in fact well above the original requirements, it is possible to do even better by using a DMA device to carry out fast memory to memory transfers without the overheads of the setups needed for the LDDR instruction.

There were two problems to be overcome with this solution. The first was the lack of memory map available to the Z80 processor, the existing hardware providing continuous RAM in only 16k pages. The second problem was the lack of fast DMA chips in the Z80 family.

The first of these problems has already been solved in chapter 2. Here a reset controlled flip-flop was used to allow alternating $\overline{\text{RD}}$ and $\overline{\text{WR}}$ signals to be sent for

certain parts of the memory map. The surface hardware simply made use of the same memory map and replaced the vertical integration hardware.

The speed of the DMA turned out to be much faster than the LDDR instruction. The predicted times of transfer can be found from analysis of the DMA cycle time. There are 388 pixels per line, 64 rows being displayed and hence 24832 bytes need to be moved to update the display once. The DMA chip, the Z8410 [76], is rated to do one transfer every 6 clock cycles and has a max clock frequency of 4MHz and hence the transfer time is given by

$$\frac{4 \times 10^6}{24832 \times 6} = 26{\cdot}85 \text{ frames s}^{-1}.$$

This corresponds to 0·0372 seconds per frame and gives an improvement of $\frac{21}{6} \times \frac{4}{6} =$ 2·33 over the LDDR command (equation 4.3). It also removes the need for software overheads in the setting up of the LDDR command.

When set to free run (no frame grab) the measured time for 200 cycles, averaged over 5 readings, was 7·70 ± 0·04 seconds. which corresponds to a frame time of 0·0385 ± 0·0002 seconds per frame.

The discrepancy, though small, was found to be due to the added time of synchronizing to the vertical blank signal and in communications delays between the PC and the Z80.

For every frame transferred it was also necessary to perform a frame grab to capture the video input signal. This takes one frame period to do, so an added time of 20ms is needed per frame. However, because the frame grab can only be initiated during the vertical blank time, the cycle would start with a frame grab, take one and a bit frame times to complete the transfer, and would then have to wait until the end of this third frame period until the next vertical blank, and hence the next frame grab, could occur. This should, theoretically result in an update time of 0·06s or 16·7 frames s^{-1}. In reality the transfer time for this hardware was found to be 16·37 ± 0·01 frames s^{-1} (averaged over 5 frames), which, taking into account the transfer times, was a very close match to the theoretical time.

One further possibility was examined at this stage. Data on the Zilog DMA device suggested that the use of a 6MHz clock was possible, though not guaranteed by the manufacturers. This was tried and a corresponding increase in performance was

observed. As this is pushing the specifications of the device, continuous operation at these speeds was not made use of.

4.4.4 Investigated modifications

Some flicker in the display could be noticed when it was being driven from the live video signal. Shielding improved the image, as did increasing the distance from the power supply where some magnetic pick up was affecting the digital rise times. Some of the flicker was also found to be reduced by shortening the cables carrying the video signals between circuit boards.

Though helping with the system, it was evident that these improvements were not sufficient to solve the picture quality problem. Two other possibilities were then looked at. One of these arises from the fact that the cycle time between bytes being read out of the display memory and the decision as to whether they are visible or are part of a hidden area, is very close to the limit of the memory access time and could cause loss of data when clock edges are not very well defined. The other possibility was that the changeover between the two banks of memory was actually becoming visible and causing a glitch on the output data lines.

The speed problem was solved by reducing the clock frequency to 1·843MHz. This now reduced the frame rate to only 50 frames a second, but did stabilize the picture and reduced the flicker to only one every few seconds. Crystals of intermediate frequency were not available, but it should be possible to go up to 70 frames per seconds without getting too close to the limit of the memory access times. It was decided that the remaining flicker was due to noise pick up in the oscilloscope trigger leads and hence the circuit was now operating as intended. Were full RF techniques to be used on a PCB version of this circuit then it should be possible to go back to the full 100Hz refresh rate.

In order to investigate the effect of the memory bank change over an input was added to the Z80 processor sensing the hidden line reset circuitry, and monitoring it after the frame update had been done. When the reset went active, the memory banks were changed over. This ensured that the beam gate was off when the memory banks were swapped over and hence any effects would not be visible. As predicted, little effect was seen, and as the time to transfer had now been increased by a small amount

in order to synchronize the signals, the modification was not left in place.

4.4.5 Future work

Clearly the most obvious addition to the wire frame generator would be to fill in the lower half of the image. As mentioned in the software section, it is possible to see the under side of some surfaces when this type of projection is used. Because of this, minimum values, as well as maximum ones, have to be stored. This would be done by a set of components similar to those used to generate the top half, but looking at the $\overline{A < B}$ instead of $\overline{A > B}$ output of the comparator. Clearly only one possible output will ever be active and hence there would be no need for extra logic and the data busses of both top and bottom circuits could be wired in parallel to the output latches.

The adaptation of these techniques to grey-scale rendering is not so readily achievable. The scanning methods available via an oscilloscope are not capable of doing this type of rendering with any degree of accuracy and hence it would be necessary to return to the normal type of raster scanned pixel format. The algorithms developed for grey-scale rendering would be easily implemented in silicon however, and, were facilities available to do this type of development, a high speed shading system should be possible.

Chapter 5

Displaying Images By Random Sequence Order Scanning

5.1 Introduction

Visual display units (VDUs) have conventionally been raster scanned. This chapter aims to look at adapting the principles of vector scanning used in chapter 4 to explore the possible improvements that such scanning can make to conventional VDUs.

Work done on this type of Human-Computer interface has shown that frame flicker in a normal computer screen will increase eye strain and can lead to mistakes of recognition [77, 78]. It is because of this that most computer companies operate their displays at frame rates greater than 70Hz, above which the human eye perceives a continuous image. Alternatives to this are to scan using Hilbert [79] or Piano [80] space filling curves, where the beam is localized to areas rather than lines and hence reduces the visibility of the flicker.

Recent work has also shown that such scanning patterns can greatly assist in real time video processing such as data compression [81, 82] and, when limited to Murray polygons [83], can be used to do half-toning and smoothing. Another method of reducing the eyes detection of flicker is to use an alternative to the sequential vertical scanning used in a normal raster display. One such method, and its applications to computer security will be the subject of this chapter.

In 1989 a patent application [84] was filed with the U.K. Patent office that outlined a method of protecting computer video signals from surveillance equipment. The image to be displayed is read out of the normal random access display memory one line at a time. Horizontally, the data is read out left to right, as normal, but the row order is not

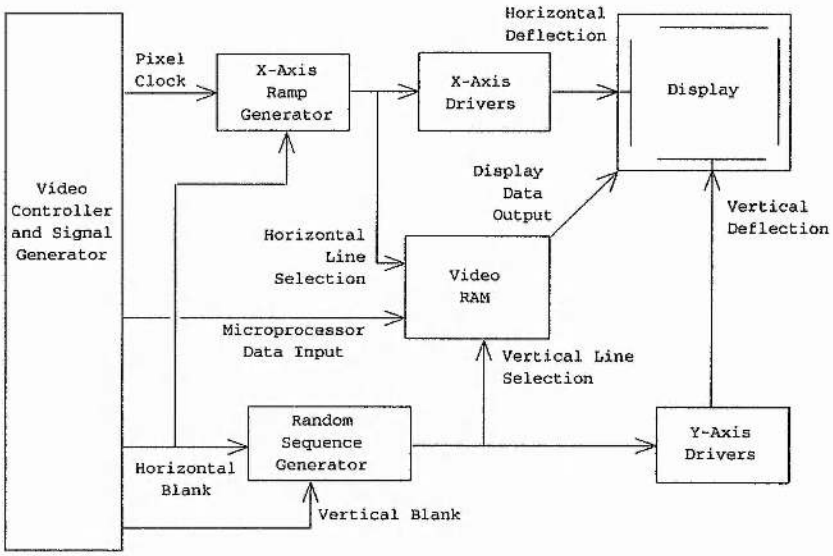


Figure 5.1: Block diagram of the VDU security patent.

sequential from top to bottom (lines 1,2,3 etc.) but is defined by a random sequence. This sequence would ideally be changed after every frame has been fully displayed. The image is recreated by sending the display device two blocks of data per line. One representing the intensity of the pixels along the line and the other describing the vertical position the line should be displayed at. The vertical position is the random sequence value that was used to select which row was to be read out.

The block diagram outlining this invention is shown in Figure 5.1.

Apart from aiding in the reduction of flicker the primary aim of this patent application is to protect displayed data from remote, clandestine sensing equipment. Video signals can be remotely detected in a number of ways. This patent aims to prevent all forms of detection that rely on the gun emissions of the display device, including normal gated magnetic or electrostatic deflection displays such as those used in IBM clones. Commercially available 'off the shelf' devices [85] allow a normal IBM VGA screen to be reproduced, in real time, from as far as 100m away. As these devices cost as little as £3000 it is easy to see why some form of cheap protection is needed if data is to be kept secure.

There are now a number of journals dedicated to providing information on computer security [86] and books have been published on the subjects of scrambling video

signals [87] and data theft [88]. All these publications have looked at preventing the interception of data while it is being transmitted rather than investigating the problems involved with the actual sending systems. Even so, they illustrate beyond a doubt that there is a huge, hidden problem to which a cheap and simple solution is urgently required.

There are only two competing protection system available today. One method is to radiate similar signals to those output from the scan coils and hence confuse the detector [89, 90] and the other, currently only available to the military, is known as *tempest*. This is based around a Faraday cage that encloses the screen, the computer generated signals and the connecting cables. In order to see the screen a set of optically transparent metal layers are used. This design is inherently bulky and at a cost of £50000 it is unlikely to be used by most PC owners. The MOD claim it is not detectable from more than 10cm away, but there are still commercial products (though in the £30000 range) that are specified to receive *tempest* protected signals at reasonably large ranges.

Preliminary market research by Marconi Electronics illustrated the need for an inexpensive protection device by the legal profession. In particular the Law Society expressed some interest in the development of a cheap video protection system, as their members deal with many sensitive issues. There are a number of standards proposed for the protection of data via software scrambling algorithms but there is, as yet, no national or international agreement on the regulation of computer hardware methods of data encryption.

An implementation of random order scanning would provide a nearly secure system. Techniques such as character recognition and continuity analysis would allow computer unscrambling of the image but this is less significant than it may seem as to work successfully surveillance must be done in real time. For normal TV signals there are 625 interlaced lines per screen and hence $625! = 6{\cdot}14 \times 10^{1477}$ combinations. Even with image analysis it would take a very large amount of computer power to manage real time unscrambling.

The are, however, two weaknesses in such a system. Firstly, it is possible to detect the vertical positioning signals in much the same way as the surveillance devices receive the intensity data, and secondly it may be possible to decode the sequence generation

circuitry and hence be able to predict the 'random' scan order.

5.2 An Electrostatic Implementation

5.2.1 Simplifications to the invention specification

For the initial prototype a reduced specification was decided upon. The use of an oscilloscope as a display device was chosen as it provided a good test bed for the digital circuitry and had the ability to display the horizontal deflections without the need for high power driver circuits which are required by the electromagnetic (rather than electrostatic) deflection mechanisms of a normal VDU. The oscilloscope also provides a very high input impedance, thus simplifying the output stages and protecting the development circuits.

Because the oscilloscope does not have built in horizontal frequency settings, relying on the input signal to provide the triggering, it is necessary to generate a trigger signal for the horizontal (x-axis) flyback. The y plates are then used to provide the vertical position of the beam. This vertical signal is changed at the end (RHS) of a line and the flyback time is used to stabilize the beam position. As the flyback time is up to one fifth of the line period (64μs for PAL systems) stabilization has to be achieved within a maximum of 12μs. Most scopes, including the one used here, are slightly under-damped, and settle within 1 cycle of their 3dB frequency, i.e. at 60MHz they settle in $\frac{1}{60\times10^6} = 17$ns. Well within requirements.

The scope used was the Hameg 60MHz dual trace unit as used previously. The scrambling circuit was to be driven by the signals available from the timing and display circuits developed in chapter 1, the intensity signal being taken directly from the most significant bit of the video data output.

In using the available video circuits, it was also necessary to further modify the specification. The existing hardware generates signal outputs for a normal raster scanned image, it is not possible to extract information from randomly selected lines. Because of this it was decided to use a fixed line scrambling order, rather than change it at the end of each frame which would be the ideal. This meant that the graphics software could create the image on the screen in a prescrambled order which, when read out sequentially, will be unscrambled by the vertical positioning circuits and will appear in

the correct vertical position. This is only possible for a 'proof of principal' circuit as, for a commercial product, the time overheads required to scramble in software would be unacceptable, particularly in a version where the random sequence is changed regularly.

One major penalty that such use of continually changing signals would incur is that the changing scan order leads to the possibility that one line may not be updated for almost 2 frame periods. This would lead to increased flicker in some instances. However, as the sequence is changing every frame, this effect should be minimal. If a full implementation were to be created further research in this area would be essential.

5.2.2 The generation of y-axis waveforms

In order to generate the vertical displacement signal, a method had to be found to map the video signal into a random sequence. By using the horizontal blank signal to clock a counter and using the vertical blank to reset it, it is possible to extract the line number, (vertical height), of the beam. If this is then fed into the address inputs of a ROM the 8 bit data output provides a programmable mapping from raster line number to scrambled line number. By feeding this into a D/A converter and a buffer stage the scope input signal can be generated.

By using only the 8 data bits of a ROM, the output would be limited to only 256 lines. One way to partially overcome this is by feeding the ninth bit output of the raster line counter into the ninth address input of the ROM and ensuring the first 56 values of the sequence contains all the numbers 1 to 56. This has the effect of providing the extra 56 lines needed to make up the 312 lines normally displayed on a PAL video monitor, simply repeating the sequence used for the first 56 lines but shifted down to lines 256 onwards. This compromise is again adequate to prove the principal.

An early choice of the mapping sequence was to scramble over only 16 lines, then repeating the sequence for lines 16 to 31 and so on. For the electrostatic implementation this can be extended to scramble over all 256 lines. The random sequence was generated by a computer program, as discussed in section 5.4.

A number of suitable D/A converters were considered. All had to be able to convert 9 bits with a settling time such that the beam, and the D/A, can be stabilized in less than the frame flyback time. Because a magnetically scanned VDU would eventually be used instead of the electrostatic oscilloscope, the faster devices were to be preferred.

Device	Speed	Price
DAC HF10BMC	25ns	NA
DAC IC10B	250ns	£15·08
DAC 610C	500ns	£10·59
DAC 7533	700ns	£5·77

Table 5.1: Datel DACs.

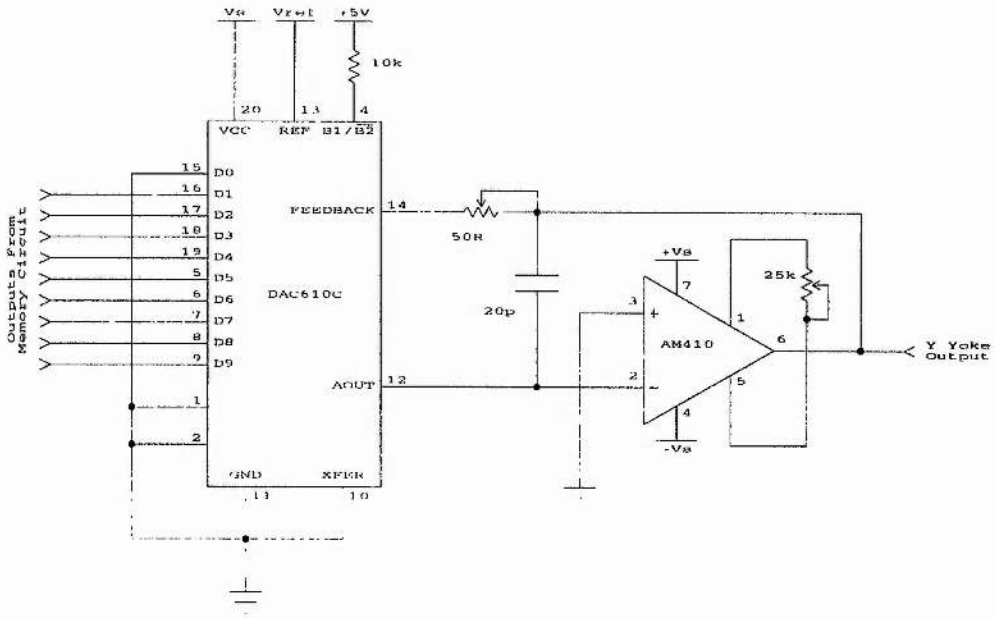


Figure 5.2: DAC610C implementation.

At the time of development the best supplier of cheap, low volume D/A converters was Datel. Table 5.1 lists some parts that were considered.

Figure 5.2 shows a test circuit used for the DAC610C [91]. Though fast enough, it relied on an obsolete amplifier, the AM410 [92]. The circuit was also found to be very sensitive to local static and overloads. Because of these problems it was decided to use a more stable device and consultation with Datel suggested the DACIC10B [93]. This was an expensive part, but it had the fastest response time and would be supported for some time to come – an important issue when considering mass production. As it is a current converter, it requires a constant current input, and a current to voltage amplifier on the output. The existing, though obsolete, AM410 amplifier was re-used as

it has a suitably high input impedance. Any similar op-amp with the same bandwidth and input characteristics could be used in any future developments.

The final circuit is shown in figure 5.3. The constant current reference source was provided by the Zener diode and fine tuning of the DC offset was provided by the trim facility on the AM410. A further buffer stage was also added to provide a gain stage, and a coarse DC offset adjust. The choice of a 2764 ROM allowed a number of different mappings to be investigated, the higher address bits being hardwired to give any desired page of ROM.

The ±15V supplies used in this circuit were developed using fixed voltage regulators similar to those used in the previous power supply circuits.

5.2.3 The generation of trigger and x-axis waveforms

Because of the need to eventually drive an electromagnetic signal, coupled with the need for a trigger circuit, it was also necessary to generate a ramp voltage for the x-axis displacement of the beam. This circuit operated in a very similar way to the vertical waveform generator, the counter was clocked by the pixel clock, and reset by the horizontal blank. In order to make a flexible prototype capable of easy modification a ROM was inserted, similar to the y-axis system. This enabled distortion correction to be made on the horizontal travel of the beam, and also gave the ability to experiment with different waveforms for the triggering of the oscilloscope version.

The bandwidth in this circuit was a problem. Video rate circuitry is required because, even though the pixel rate is only 8MHz, the actual pixel positions need to be settled as fast as possible to avoid distortions and provide sharp edges for the triggering. To this end the PNA7518 [5], used previously, was an ideal choice. It provides sufficient bits of conversion (necessary to allow 300 or more pixels per line) and has a bandwidth of 30MHz. This was coupled with an NE5532 video op-amp to provide the driving voltages.

Figure 5.4 shows the final version of the beam deflecting circuit, incorporating the PNA7518.

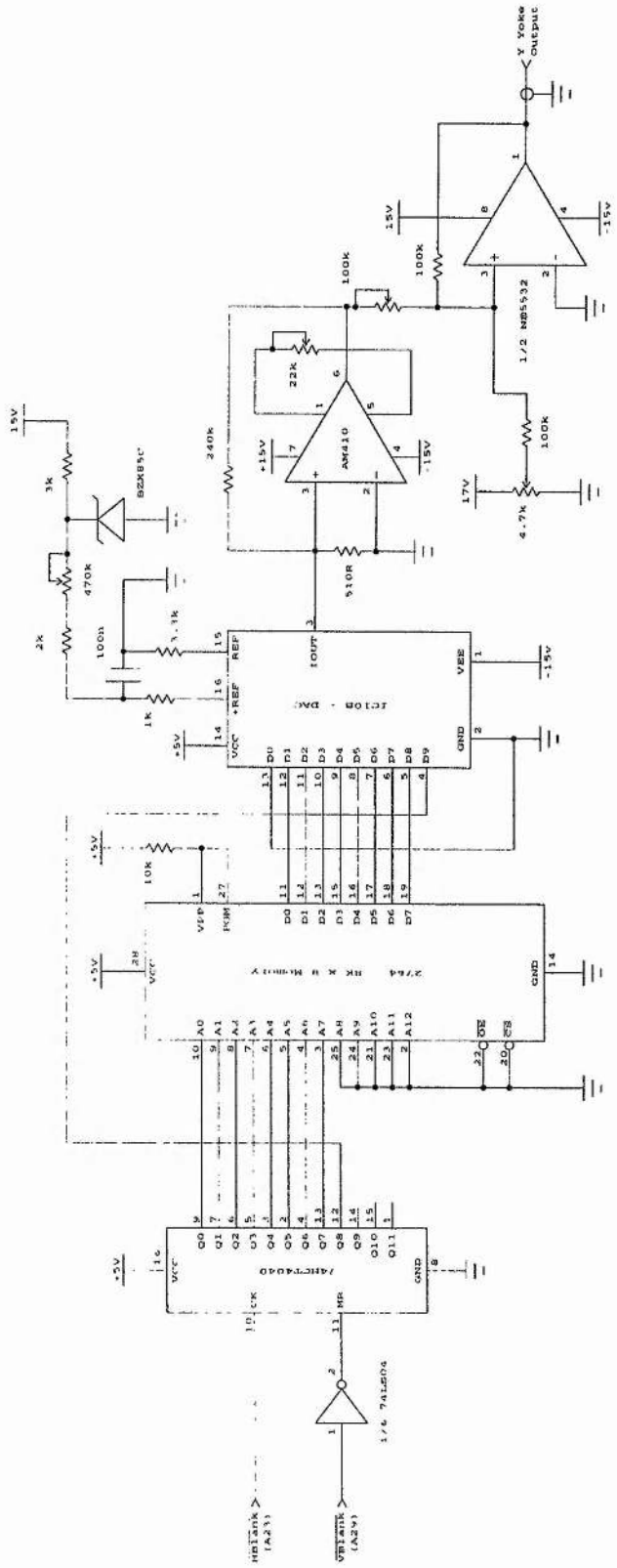


Figure 5.3: The final circuit for y-axis generation.

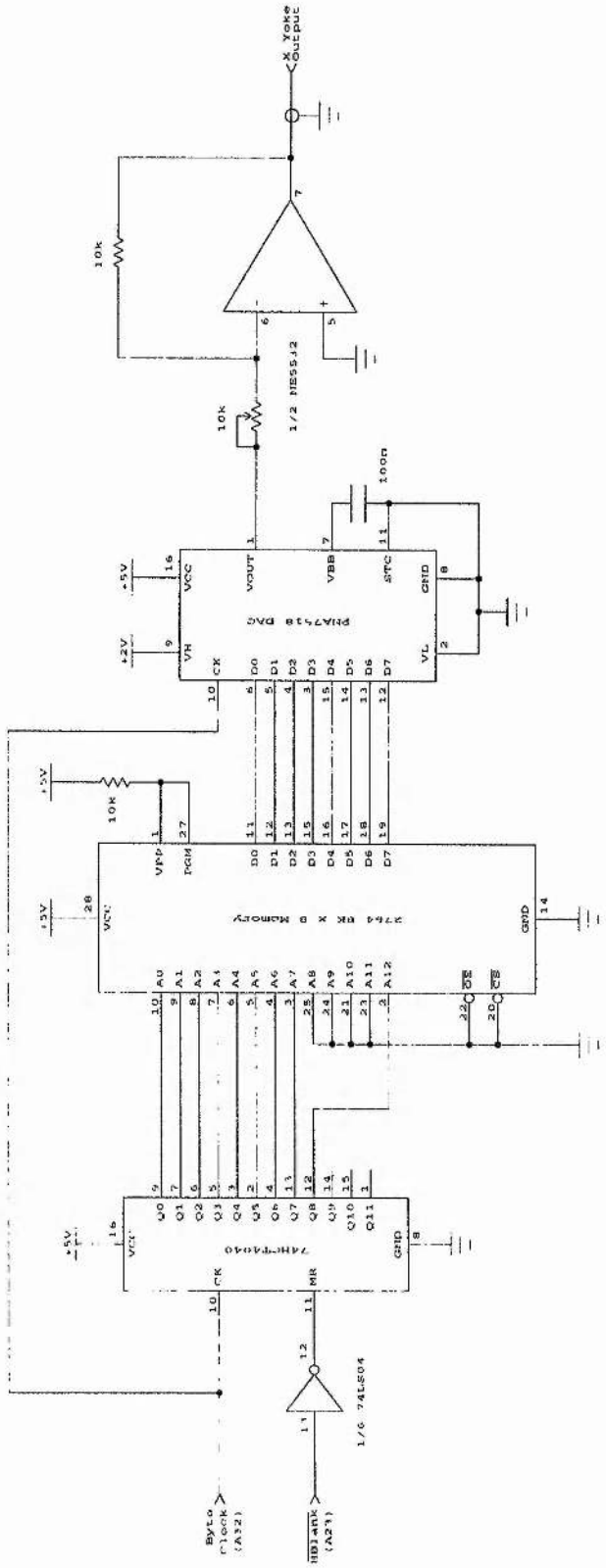


Figure 5.4: x-axis deflection circuit.

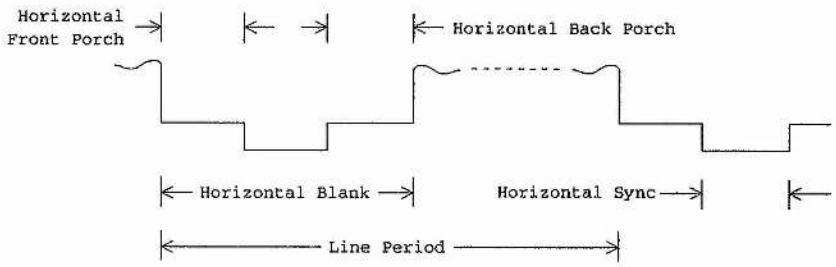


Figure 5.5: Horizontal control timings.

5.2.4 Improvements to the original specifications

Two methods of improving the invention were researched. Firstly, the circuitry driving the y-axis had to have an adequate bandwidth to settle the beam correctly even when it is moved from the bottom line to the top line in one flyback time. This means that there is no need for control signals such as the vertical synchronization, vertical blank and front/back porches. The removal of these means that a detection circuit that extracts the vertical signal from the beam would not be able to lock on to the frame frequency. The display driver of such a surveillance device would lose the vertical hold and the picture would scroll around from top to bottom.

This idea can be extended to remove part of the front and back porches and all of the synchronization signal from the horizontal blank time. This is possible, on an oscilloscope version, because the flyback time is greater than it is on a normal electromagnetic coil. Provided the blank period is long enough for the flyback and the beam position settling to occur it can be reduced to beyond the point where a normal system could lock onto it. Figure 5.5 shows the composite signal and where the various timing points are.

The second improvement was to vary the number of lines displayed in each frame. By removing the vertical synchronization signal, the concept of a frame has effectively been removed. This means that, provided flicker does not creep in, some lines can be missed out of some frames, some can be displayed more than once, or the blank periods can be randomly extended. All this would make any computer decoding much harder and would ensure that much more flexible (and hence expensive) circuitry would be needed to even correctly detect the scrambled signal let alone unscramble it.

5.2.5 Results

Once a prototype had been designed and the construction verified, an emissions test was done by the Marconi Signals laboratory based at Edinburgh. The prototype circuitry was set up in a shielded lab, along with a number of aerials that, between them, covered most of the detectable radio frequencies. For test purposes, a standard monitor was driven with a normal image. This enabled the detection circuitry to find which frequencies were being emitted. The most obvious one was the pixel rate (8MHz) which had the intensity signal modulated onto it. Large emission intensities were also detected at the line rate, the PC processor clock rate (12Mhz) and the base frequency of the drivers which is three times the pixel rate (25MHz). There were also some intensities detected at lower rates, in particular the 50Hz mains frequency. Interestingly, some of the strongest signals were actually detected via harmonics of the pixel rate, some of which were in the GHz region. These strongest signals were easily detected and the displayed text was easily read on the detector screen, clearly illustrating the ease with which remote detection can be done.

The next stage was to scramble the display and connect up the unscrambling circuit. This produced little change to the pattern of detected frequencies. The normal display was now turned off and the resultant emissions from just the scrambling circuitry were analysed.

What was seen was exactly as predicted. There were a number of detectable frequencies some of which gave a very clear, but *scrambled* image. The mapping tested was using a full 256 line scramble and a number of possible pictures were analysed. Figure 5.6 show some text being displayed on the monitor, this is a version of what was seen by the Marconi test equipment, before the synchronization signals were removed. Figure 5.7 shows the unscrambled version that was displayed on the oscilloscope screen.

Once the synchronization modification, as described in section 5.2.4, were made the detected picture broke up. There was clearly something there, but the available adjustment to the Marconi scanning circuits was not able to find the line rate and hence the picture scrolled round vertically and jittered horizontally. Given time it may have been possible to make a locking circuit for the scanning equipment but, as there was over £100000 worth of detection equipment being used, it was clear that this scrambling device would defeat any small time criminal activities.

Figure 5.6: The scrambled text as seen by Marconi.

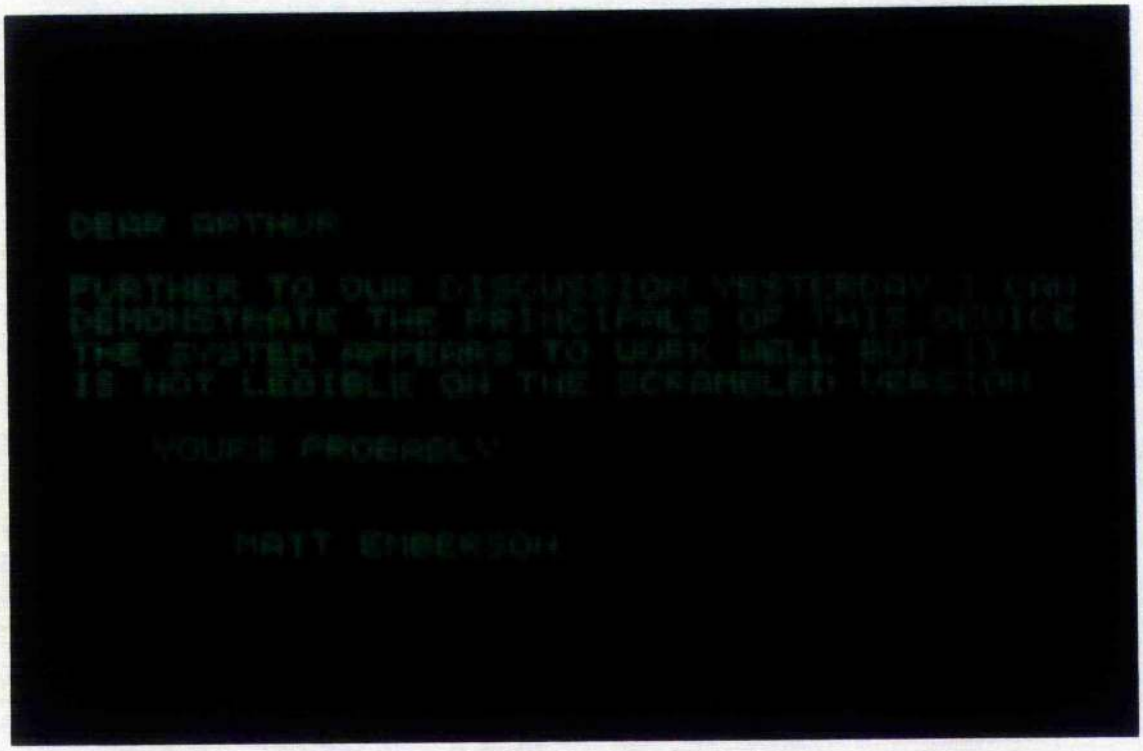


Figure 5.7: The readable version of the text.

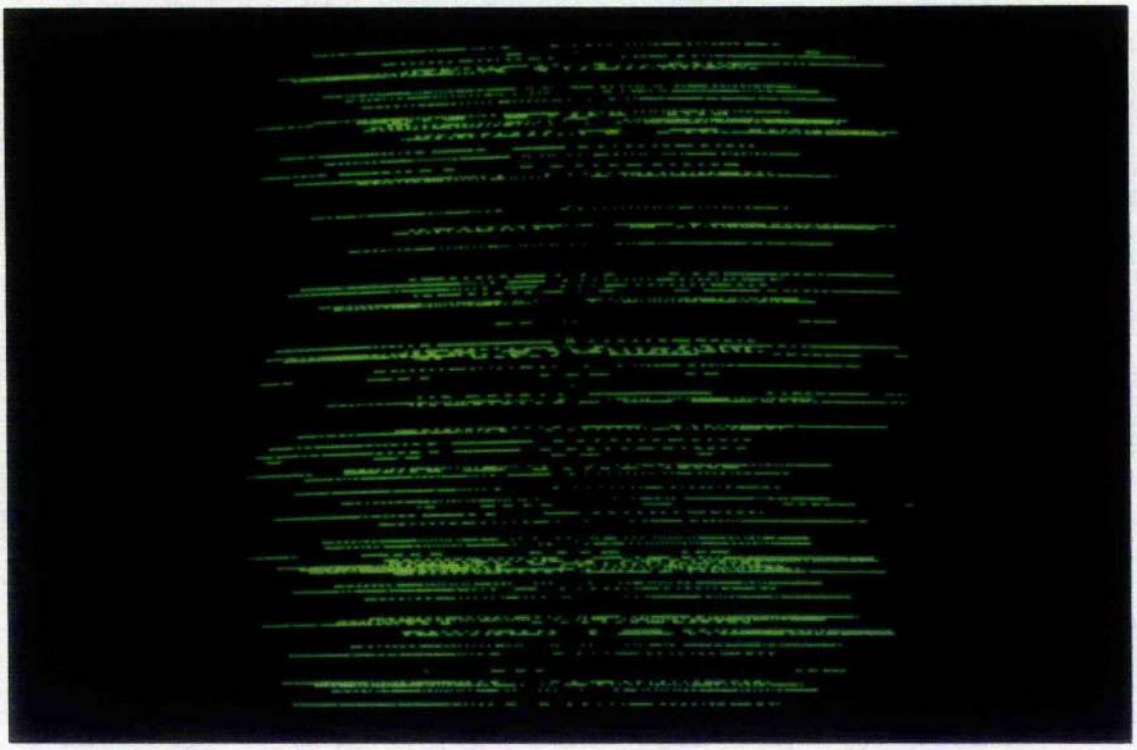


Figure 5.8: A black and white scrambled graphics image.

Figures 5.8 and 5.9 show the results of the same tests but using a graphics image rather than text.

Both these two figures, and the previous two, only used one bit intensity output, as a variably gateable display was not available. Figure 5.10 shows the scrambled version of the grey-scale image shown in figure 4.10. These are both displayed on the monitor but serve to illustrate the type of effect a full grey-scale version would achieve.

5.3 Investigation of Electromagnetic Deflection

5.3.1 Existing electromagnetic displays

Once the prototype had been shown to work well using electrostatic deflection, the next stage was to investigate the modifications needed to use a normal, magnetically scanned monitor. For computer applications the industry standard is now the IBM VGA format. VGA screens use a three colour input, with the synchronization signals being provided by a fourth input containing composite sync data. The decoding circuits simply remove the vertical and horizontal synchronization pulses and use them to lock in the vertical and horizontal driver frequencies. These drivers are complex, particularly in older designs,

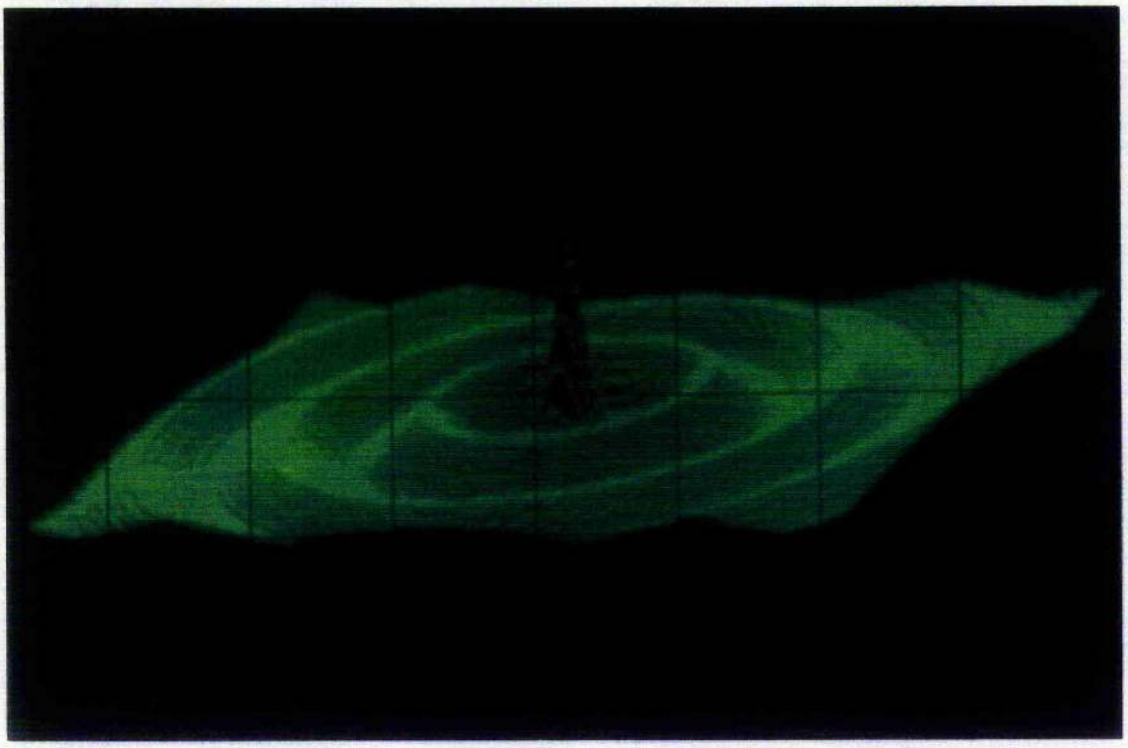


Figure 5.9: The unscrambled black and white graphics image.

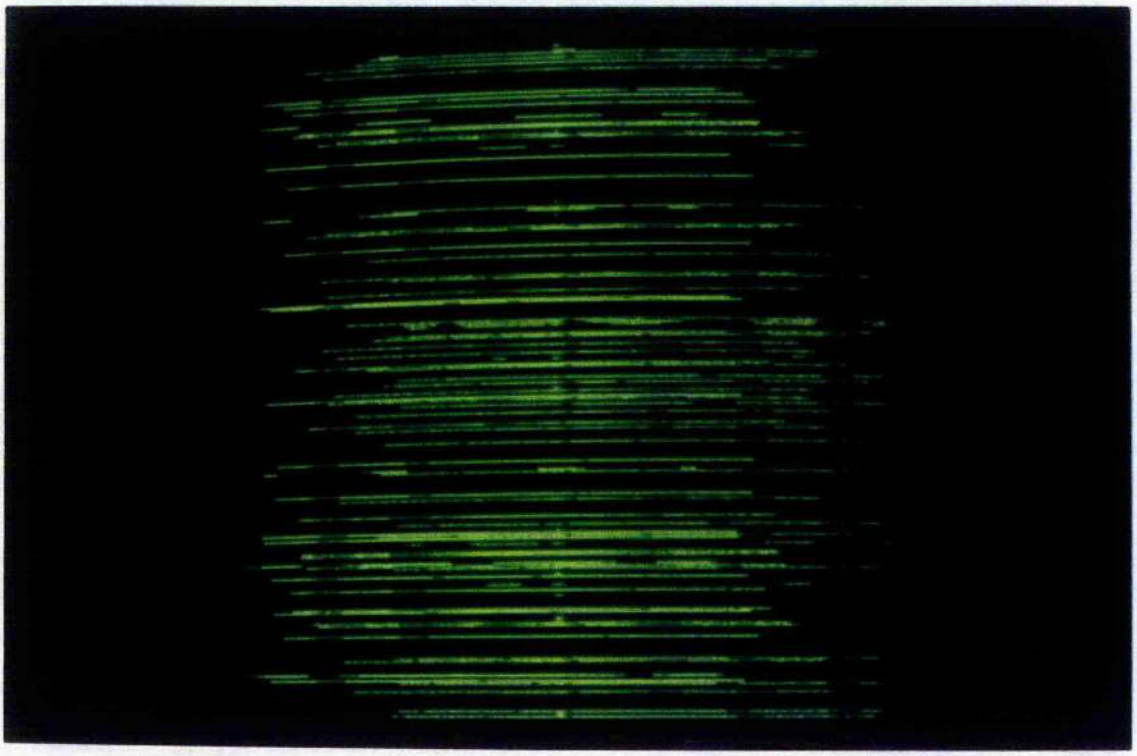


Figure 5.10: A grey-scale scrambled image.

where the beam deflection distance is not a linear function of the voltages produced by the drivers. For test purposes this is not a problem but, as will be discussed later, it would be a much more significant factor when building to production standards.

The monitor chosen for testing was a simple 10 inch closed circuit unit, type TVM-10. This is a typical VDU unit and was preferred only because circuit diagrams and board layouts were available.

5.3.2 Trials with an existing monitor

The first approach was to examine the built in drivers of the VDU and attempt to use the high current stages directly. The driving stage for the vertical beam deflection was found to be a simple push pull, resonant current amplifier. Its inputs were filtered to remove RF interference and any DC component and the coil was used in parallel with a large capacitor and resistor to resonate at the required frame frequency. In most monitor designs the input waveform is also modified by the final stage circuitry to correct it for non linearities in the deflection system [94]. This is particularly noticeable in the horizontal axis deflection where the yoke voltage would normally be linearly increasing but, as the deflection is an angular displacement, corrections are needed for the edges of the screen. As the linearity problem is less severe in the vertical movement for this particular monitor, the vertical ramp input was found to be only slightly modified by the output stages. Clearly this is a problem specific to this make of monitor but it is likely that similar problems would be encountered with all forms of magnetic deflection.

Figure 5.11 shows the input waveform that drives the vertical push-pull stage, as can be seen it is slightly discontinuous in the centre of the sweep but a linear ramp can still be used as an approximating waveform.

By removing capacitors from the input circuit it was hoped to remove the filtering and so enable direct drive of the beam. The amplifier inputs were disconnected from the drive circuitry and a square wave was fed into them instead. By adjusting the frequency of the input signal to a multiple of the horizontal scan rate, the square wave appeared as a triangular waveform on the display screen. This is because the current, and therefore the beam, rise time was not fast enough to position the beam correctly, as defined by the square wave voltage, within the half period of the input square wave.

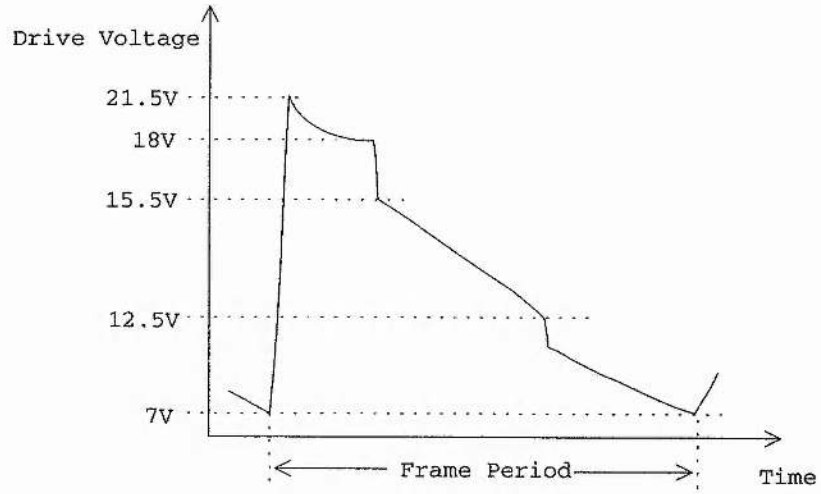


Figure 5.11: Existing inputs to the vertical coil drivers.

This meant that while the square wave was in its positive cycle the beam was rising up the screen whereas when the square wave went negative, the beam started to sink down, never actually reaching its specified position.

When the frequency was increased to the next multiple of the line scan frequency it was found that the deflection of the beam was smaller because there was less time for the beam to get to the required position. It was also found that if the voltage was increased then the rate of deflection was greater. Clearly a trade off exists. By using higher voltages, it is possible to reduce the current rise time in the coil, but at some point either the coil or the amplifier will break down due to excessive voltage levels. The conclusion had to be that the existing drivers were not capable of handling the rise times that are required to generate a full scale deflection of the beam within the flyback time.

5.3.3 Software correcting the display

Three further developments were now considered. Since some beam deflection could be achieved it was possible to scramble in such a way that the beam never had to go more than say 16 line widths during any one flyback. This was the maximum deflection that could be stabilized using the existing hardware.

The second development would be to incorporate, an intelligent controller into the

Figure 5.12: Text scrambled over 16 lines.

drivers that would specify a deflection of say 40 lines when only a 20 line deflection was needed. This would increase the voltage, and hence the rise time. When the beam got to the required position, the voltage offset would be returned to the correct value, and the beam stabilized. The latter method is best implemented by a current amplifier, and will be dealt with later.

Limiting the beam deflection has a very large effect on the scrambling. The number of permutations, p is given by

$$p = n!\,(t \text{ div } n) + (t \bmod n)!,$$

where n is the number of lines scrambled over (in this case 16), t is the total number of lines on the screen and the 'div' and 'mod' are the integer division and remainder functions respectively.

For 16 line scrambling and a screen of 311 lines (as used by the test hardware in non-interlaced mode) this gives only $3{\cdot}98{\times}10^{14}$ permutations. Experiments done with scrambling over 16 lines worked quite well, but, as the human eye is very good at character recognition even when the characters are not perfect, it was often possible to interpret the images. This was confirmed on the electrostatic prototype where some text examples was almost readable. One such example is given in figure 5.12 with the original text given in figure 5.13. The scrambled text, though not readable, shows some structure and the presence of spaces in the text is clearly seen. It is also possible to simply reconstruct letters where there are no interactions with other characters. The 'H' in 'HUMAN' and the 'T' in 'THIS' are good examples.

The third development was only possible because a fixed scrambling pattern and a known hardware setup were being used. Firstly a scrambling pattern could be programmed into the ROM, and an image displayed as if no scrambling had been used. This would give a picture, on the monitor, that represented the unscrambling algorithm

Figure 5.13: Example unscrambled text.

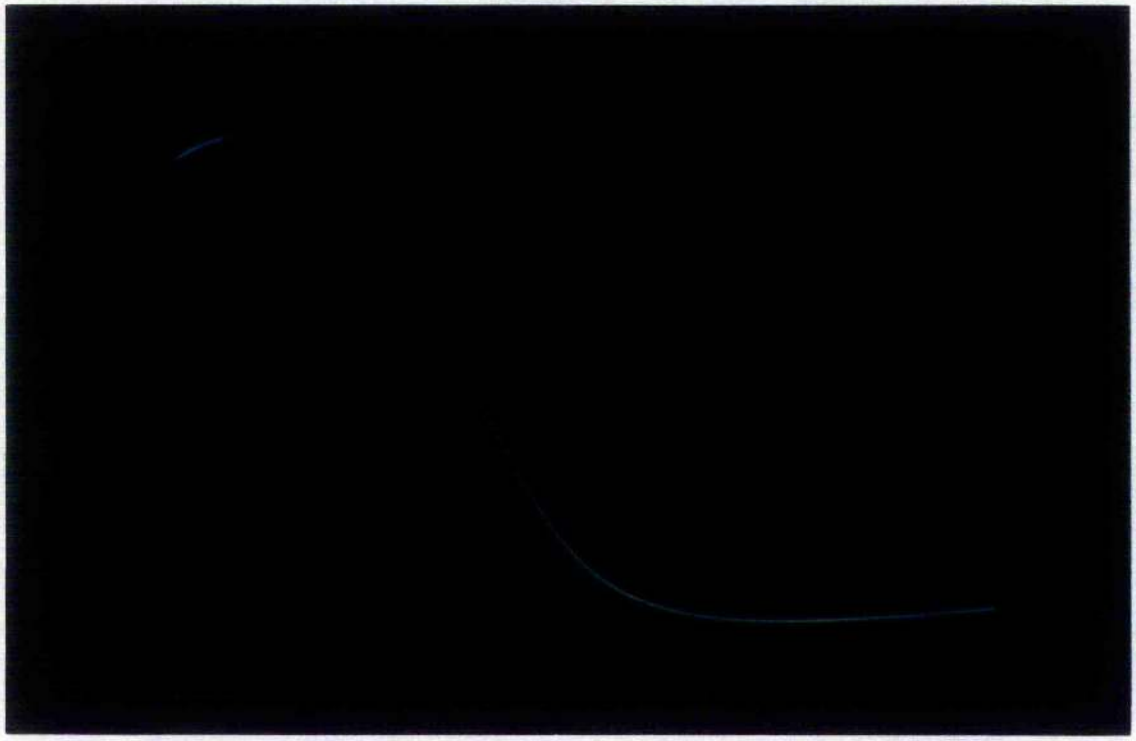


Figure 5.14: The effects of scrambling a diagonal line.

as the hardware operation assumes that the picture has been drawn in its scrambled version. If a diagonal line is drawn at 45 degrees, it is now possible to find out what the unscrambling algorithm actually is. In a perfect system this would be identical to the pattern programmed into the ROM but, as there is a limited rise time, the actual unscrambled picture is quite different. An example output from such a scrambled line, on an electromagnetic display is shown in figure 5.14. This has been programmed with a repeating sequence of 16 different scrambled values. The effects are quite good at the start of the sweep, but as the beam goes further down the screen, filtering effects start to remove the small changes and they soon vanish completely.

By manually working out the order in which the lines as displayed, it is possible to set up the software to use this 'correct' scrambling order. To assist in the analysis of

Line Number	Expected Position	Actual Position
0	0	1
1	3	9
2	11	5
3	13	12
4	14	3
5	10	6
6	5	4
7	1	11
8	6	7
9	12	2
10	7	13
11	8	0
12	4	8
13	2	14
14	9	10
15	15	15

Table 5.2: Visually determined line ordering for 16 line scrambling.

the screen, a program was written that displayed the diagonal line at half intensity, but allowed selection of one horizontal line of pixels to be displayed at full brightness. This program was then used to compile the mapping shown in table 5.2. In this example only 16 lines were scrambled over to simplify the analysis.

The biggest problem with this method of scrambling is that the beam is not stable while it is visible. The amplifiers will continue to move the beam vertically after the gun has started displaying the line. This means that the above set of data, while correct for the points used, may not be correct for all horizontal points on any given line. This could be corrected for if all the horizontal positions for every point were determined, but as nearly 10 thousand points would need to be plotted and used to map backwards, this solution is clearly not acceptable.

It can also be seen that, though the points are evenly spaced along their y-coordinate (which, if the beam is still rising, may well not be horizontal), they are not evenly spaced vertically. It might be possible to stabilize the beam at the start of each line, thus making all lines horizontal, but the complexity needed to ensure that this resulted in lines that were evenly spaced vertically would be unacceptable.

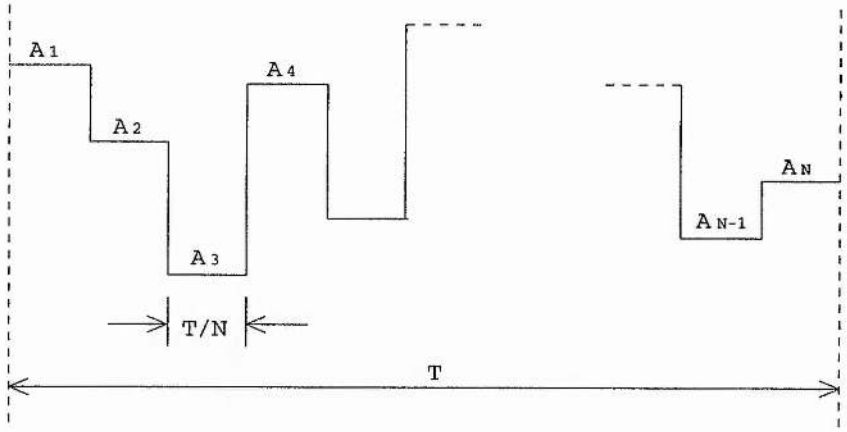


Figure 5.15: The general form of the y-axis waveform.

5.3.4 Fourier analysis of y-axis waveforms

In order to fully understand the driver circuitry, it was decided to look into the waveforms in more detail. The aim was to find the frequency components that had to be present and to see how the waveforms were affected by limited bandwidth. By Fourier Transform theory, coupled with the fact that it is possible to view the sequence of voltages used to drive the vertical coils as an infinitely repeating waveform, the drive signal can be expressed in the frequency domain rather than time domain.

Figure 5.15 represents an arbitrary form of the driving waveform where the A_i represent the amplitudes of the deflection waveform for each line period. For such a waveform te Fourier Series equations can be shown as

$$s(t) = \sum_{n=-\infty}^{+\infty} \beta_n e^{jn\omega t},$$

$$\beta_n = \frac{1}{T}\int_0^T s(t)e^{-jn\omega t}dt. \tag{5.1}$$

For the amplifier stages the amplitude spectrum is given by

$$|\beta_n| = \sqrt{\Re e^2(\beta_n) + \Im m^2(\beta_n)}.$$

In this case equation 5.1 can be solved to get

$$\beta_n = \frac{1}{T}\sum_{p=1}^{N}\int_{(p-1)\frac{T}{N}}^{p\frac{T}{N}} A_p e^{-jn\omega t}dt,$$

$$= \frac{1}{T}\sum_{p=1}^{N}\left[\frac{A_p e^{-jn\omega t}}{-jn\omega}\right]_{(p-1)\frac{T}{N}}^{p\frac{T}{N}},$$

$$= \frac{1}{T}\sum_{p=1}^{N}\frac{A_p}{-jn\omega}\left[e^{-jn\omega p\frac{T}{N}} - e^{-jn\omega(p-1)\frac{T}{N}}\right]. \tag{5.2}$$

Now since

$$\omega = 2\pi f = \frac{2\pi}{T},$$

the amplitude spectrum is given by

$$|\beta_n| = \left|\frac{1}{-jn2\pi}\right|\left|\sum_{p=1}^{N} A_p e^{-jnp\frac{2\pi}{N}}\left[1 - e^{jn\frac{2\pi}{N}}\right]\right|.$$

This can be simplified to get

$$|\beta_n| = \frac{1}{2\pi n}\left|1 - e^{jn\frac{2\pi}{N}}\right|\left|\sum_{p=1}^{N} A_p e^{-jnp\frac{2\pi}{N}}\right|,$$

$$= \frac{1}{2\pi n}\left|1 - \cos\left(\frac{2\pi n}{N}\right) - j\sin\left(\frac{2\pi n}{N}\right)\right|\left|\sum_{p=1}^{N} A_p e^{-jnp\frac{2\pi}{N}}\right|,$$

$$= \frac{1}{2\pi n}\sqrt{\left(1 - \cos\left(\frac{2\pi n}{N}\right)^2\right) + \sin^2\left(\frac{2\pi n}{N}\right)}\left|\sum_{p=1}^{N} A_p e^{-jnp\frac{2\pi}{N}}\right|,$$

$$= \frac{1}{2\pi n}\sqrt{1 - 2\cos\left(\frac{2\pi n}{N}\right) + \cos^2 + \sin^2}\left|\sum_{p=1}^{N} A_p e^{-jnp\frac{2\pi}{N}}\right|,$$

$$= \frac{1}{2\pi n}\sqrt{\left(2 - 2\cos\left(\frac{2\pi n}{N}\right)\right)}\left|\sum_{p=1}^{N} A_p e^{-jnp\frac{2\pi}{N}}\right|, \tag{5.3}$$

$$= \frac{1}{2\pi n}\sqrt{\left(2 - 2\cos\left(\frac{2\pi n}{N}\right)\right)}\left|\sum_{p=1}^{N} A_p \cos\left(\frac{2\pi pn}{N}\right) - j\sum_{p=1}^{N} A_p \sin\left(\frac{2\pi pn}{N}\right)\right|.$$

Again, using the identity

$$|x| \equiv \sqrt{\Re e^2(x) + \Im m^2(x)},$$

it can be shown that

$$
\begin{aligned}
|\beta_n| &= \frac{1}{2\pi n}\sqrt{\left(2-2\cos\left(\frac{2\pi n}{N}\right)\right)}\times \\
&\quad \sqrt{\left[\sum_{p=1}^{N} A_p \cos\left(\frac{2\pi pn}{N}\right)\right]^2+\left[\sum_{p=1}^{N} A_p \sin\left(\frac{2\pi pn}{N}\right)\right]^2}, \qquad (5.4)\\
&= \frac{1}{2\pi n}\sqrt{\left(2-2\cos\left(\frac{2\pi n}{N}\right)\right)}\times \\
&\quad \sqrt{\sum_{p=1}^{N}\sum_{q=1}^{N} A_pA_q\left[\cos\left(\frac{2\pi pn}{N}\right)\cos\left(\frac{2\pi qn}{N}\right)+\sin\left(\frac{2\pi pn}{N}\right)\sin\left(\frac{2\pi qn}{N}\right)\right]},\\
&= \frac{1}{2\pi n}\sqrt{\left(2-2\cos\left(\frac{2\pi n}{N}\right)\right)}\sqrt{\sum_{p=1}^{N}\sum_{q=1}^{N} A_pA_q\cos\left(\frac{2\pi n}{N}(p-q)\right)},\\
&= \frac{1}{2\pi n}\sqrt{\left(2-2\cos\left(\frac{2\pi n}{N}\right)\right)}\times \\
&\quad \sqrt{\sum_{p=1}^{N} A_p^2+2\sum_{p=1}^{N-1}\sum_{q=p+1}^{N} A_pA_q\cos\left(\frac{2\pi n}{N}(p-q)\right)}. \qquad (5.5)
\end{aligned}
$$

Which is as far as it was possible to get analytically. However, since

$$
\begin{aligned}
\left|\sum_{p=1}^{N} A_p e^{-jnp\frac{2\pi}{N}}\right| &= \left|\sum_{p=1}^{N} A_p\left[\cos\left(\frac{2\pi pn}{N}\right)-i\sin\left(\frac{2\pi pn}{N}\right)\right]\right|,\\
&\leq \sum_{p=1}^{N}|A_p|\left|\cos\left(\frac{2\pi pn}{N}\right)-i\sin\left(\frac{2\pi pn}{N}\right)\right|
\end{aligned}
$$

and

$$
\begin{aligned}
|\cos x - i\sin x| &= \left(\cos^2 x+\sin^2 x\right)^{\frac{1}{2}},\\
&= 1,
\end{aligned}
$$

then

$$
\left|\sum_{p=1}^{N} A_p e^{-jnp\frac{2\pi}{N}}\right| \leq \sum_{p=1}^{N}|A_p|.
$$

By putting this back into equation 5.3 the result is

$$|\beta_n| \leq \frac{\sum_{p=1}^{N} |A_p|}{2\pi n} \sqrt{2 - 2\cos\left(\frac{2\pi n}{N}\right)}.$$

By computer analysing this envelope a plot of the maximum amplitude spectrum for this waveform can be created. Such a plot is shown in figure 5.16 along with the amplitude spectrum of the waveform used to generate the full 256 line scrambling demonstrated earlier. As can be seen the required bandwidth for an amplifier being used on this example is as large as the envelope, since all frequencies appear at approximately the same amplitude. Reduction of the bandwidth would lead to loss of the higher frequency components and a corresponding reduction in the signal accuracy.

The final analytical result, equation 5.5, is not suitable for computational purposes because if has a nested sum making it $O\{N^2\}$ (of order N^2). Because of this the program used equation 5.4 as this is only $O\{N\}$.

One interesting observation that can be made from this data is that it is periodic, though not necessarily symmetric, within one period. Mathematically equation 5.2 gives

$$\begin{aligned} \beta_n &= \frac{1}{-jn2\pi} \sum_{p=1}^{N} A_p e^{-jpn\frac{2\pi}{N}} \left[1 - e^{jn\frac{2\pi}{N}}\right], \\ &= \frac{1}{-jn2\pi} \left[1 - e^{jn\frac{2\pi}{N}}\right] \sum_{p=1}^{N} A_p e^{-jpn\frac{2\pi}{N}}. \end{aligned} \tag{5.6}$$

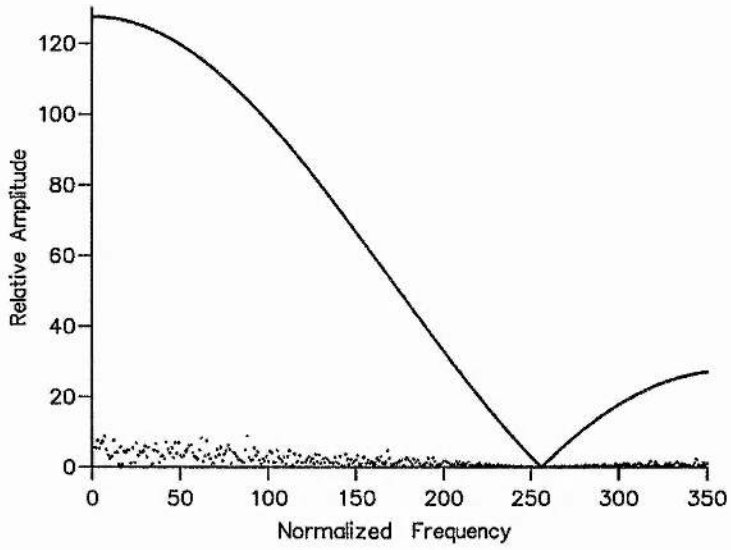
An analysis of β_{n+N} leads to

$$\begin{aligned} \beta_{n+N} &= \frac{1}{-j(n+N)2\pi} \left[1 - e^{j(n+N)\frac{2\pi}{N}}\right] \sum_{p=1}^{N} A_p e^{-jp(n+N)\frac{2\pi}{N}}, \\ &= \frac{1}{-j(n+N)2\pi} \left[1 - e^{jn\frac{2\pi}{N}} e^{j2\pi}\right] \sum_{p=1}^{N} A_p e^{-jpn\frac{2\pi}{N}} e^{-jp2\pi} \end{aligned} \tag{5.7}$$

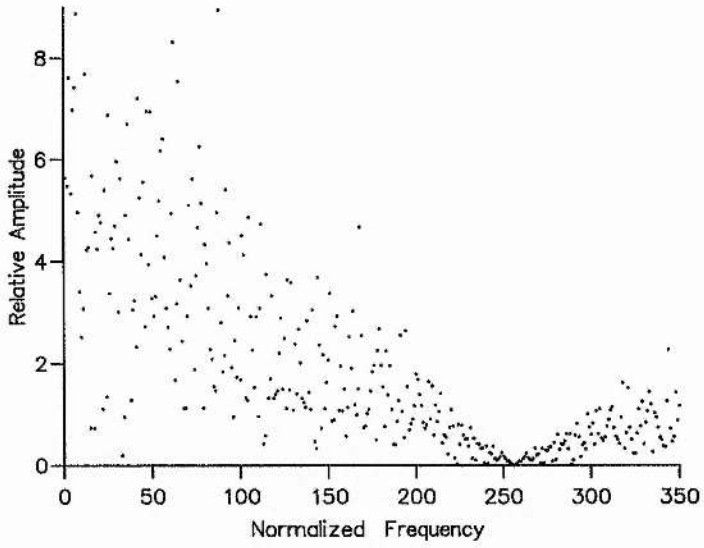
and since $e^{j2\pi} = 1$ equations 5.6 and 5.7 can be combined to get

$$\beta_{n+N} = \frac{n}{n+N}\beta_n.$$

i.e. the spectrum repeats but diminishes each cycle by an amount $\frac{n}{(n+N)}$.



(a) Fourier waveform and envelope.



(b) Magnified waveform.

Figure 5.16: The amplitude spectrum envelope and an example waveform.

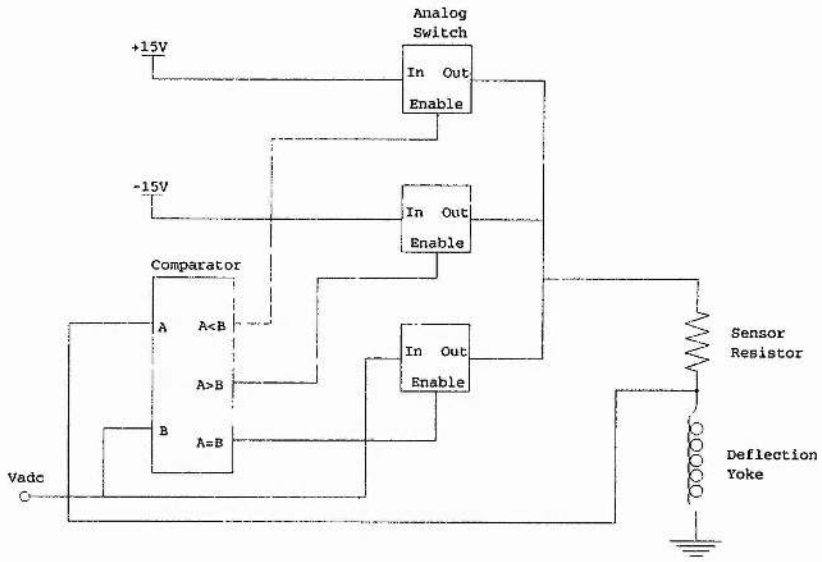


Figure 5.17: A proposed schematic for a supply rail driver.

5.3.5 Hardware implementations

As mentioned previously the required driving stage for the electromagnetic screen would have to be one that gives a very small current rise time in the deflection yoke. Clearly the fastest rise-time would be when the coil is driven from the supply rail.

Figure 5.17 shows a suggested schematic for such a driver. There are, however, a number of problems with this circuit. Firstly, the stabilizing effect of feedback is not available to the driver and hence it will suffer very strongly from overshoot and will need extensive damping in order to make it stable. The second problem, and the one that prevented this from being developed further, is that calculation showed that switches capable of reacting fast enough, and also capable of passing the large currents required, are simply not available. The use of FET switches would be ideal, but then the signals would have to be combined in a summing amplifier system with a resulting loss of the benefit of the directly driven output stage.

Because of this it was necessary to develop a current amplifier that could drive the input stages correctly. Two such circuits were investigated, the first is shown in figure 5.18 and the second, the Howland current source, is shown in figure 5.19.

The Howland has the big advantage of being able to drive the coil direct to ground. In a retrofit circuit, this would greatly simplify the modifications needed to existing

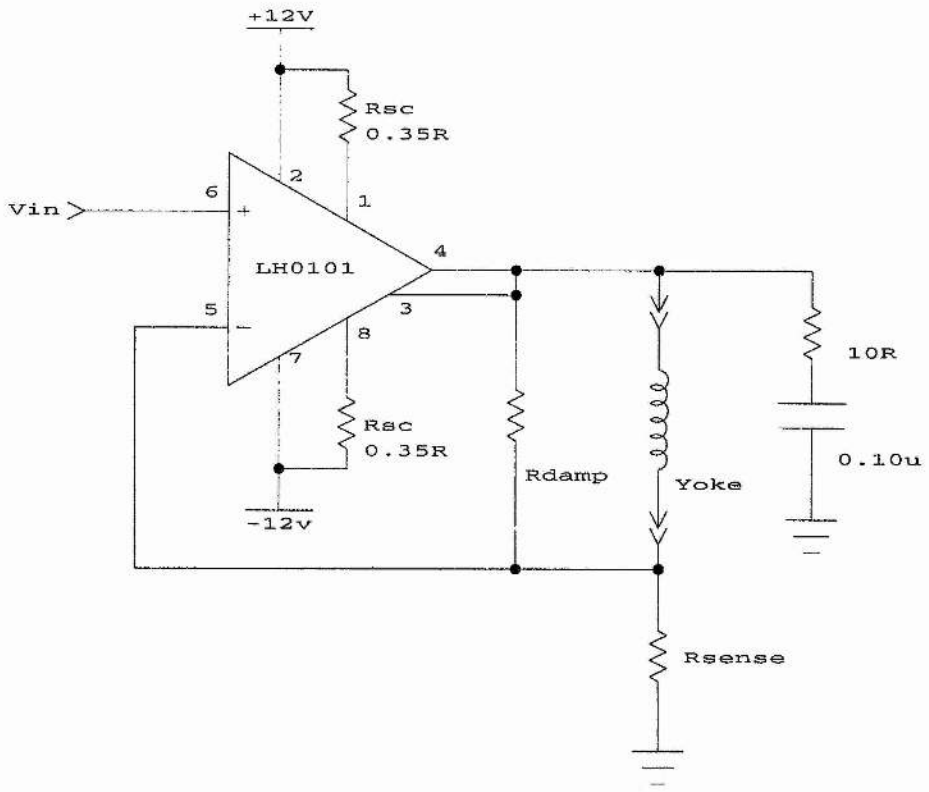


Figure 5.18: A traditional current amplifier.

circuitry, but has the disadvantage that to get good stability the resistors must be very closely matched and this would usually call for precision tolerances. It was found that the Howland also gave worse rise times for the same amplifier. This, combined with the price premium of precision resistors, forced the decision to adopt the conventional amplifier for the prototype circuitry.

Theory states that, for this amplifier,

$$\text{Short Circuit Limit} \sim \frac{0{\cdot}6}{R_{sc}},$$

$$I_{\text{yoke}} = \frac{V_{\text{in}}}{R_{\text{sense}}}. \tag{5.8}$$

By analysing the scan coils, R_{yoke} is found to be about 10Ω, and that a half scale deflection could be obtained by putting 20V DC across the coil. The amplification part of this signal is not needed here as there is ample output capability from the voltage

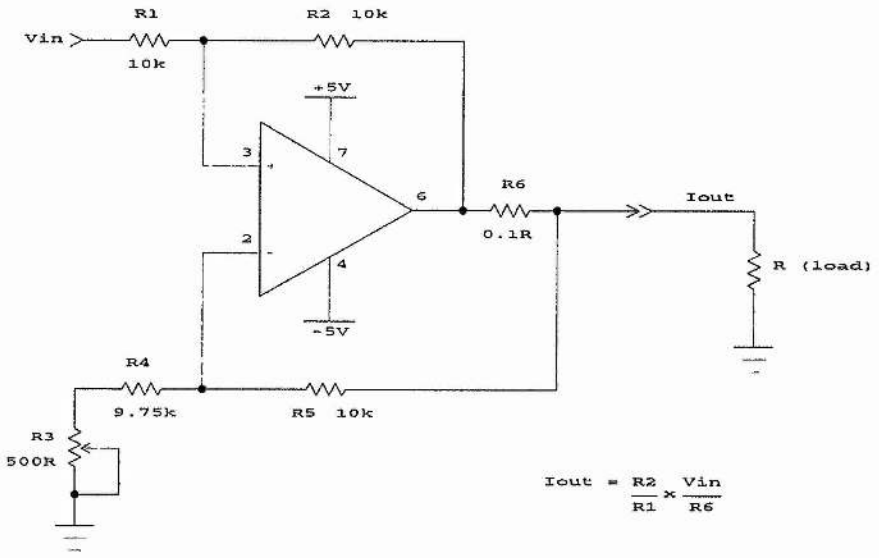


Figure 5.19: The Howland current source.

sources developed earlier. Combining this data with equation 5.8, it can be shown that $R_{\text{sense}} = 2{\cdot}5\Omega$. This leads to a power rating of 8W.

The next decision that had to be taken was which amplifier to use. High current handling was essential, but there are a wide range of parameters that could affect the output, including slew rate and power bandwidth.

In order to evaluate the amplifiers ability to drive the coils they were fed from a square wave source. The output on the screen appeared as a set of diagonal lines (see section 5.3.2) which were then used to evaluate the rise time by measuring the angle of the lines relative to the horizontal. The larger the angle, the lower the rise time and hence the better the amplifier. Because the coil was not perfectly horizontal, it was also necessary to take into account the angle of the beam when no vertical deflection was present. Figure 5.20 shows this layout.

Given the period, T, of the square wave and L the length of one half cycle of this square wave as measured on the screen, then the rise time in centimeters per μs can be found. This can then be used to find rise times for a square wave of any frequency. The rate of displacement, R, is given by

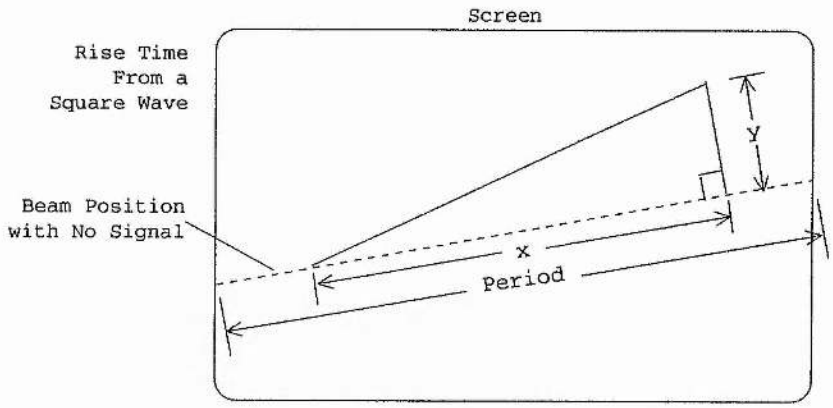


Figure 5.20: Measurement of the rise time of the amplifier.

Device	Bandwidth	I_{max}	V_{supply}	Rise Time	No. Lines
EL2008CT	55MHz	1A	$\pm$15V	$<$0·03cmμs^{-1}	7·5
LM12C	300KHz		$\pm$25V	0·067cmμs^{-1}	16·8
			$\pm$15V	0·037cmμs^{-1}	9·3
			$\pm$12V	0·026cmμs^{-1}	6·5
MOS248			$\pm$55V	0·074cmμs^{-1}	18·5

Table 5.3: Results and basic parameters of some amplifiers.

$$R = \frac{y2L}{xT}.$$

Table 5.3 show the results obtained for a selection of amplifiers and buffers, along with their basic parameters. The input square wave frequency was 16KHz, fed at 1·24V peak to peak. The amplifiers were chosen because they illustrate a good cross-section of the desired parameters.

Also included in table 5.3 are the number of lines each amplifier could scramble in a flyback time of 12μs. The calculation is based on a screen height of 14·4cm and 312 lines per screen. In order to scramble the entire screen a rise time of $\frac{14·4}{12\times10^{-6}}$=1·2cm$\mu s^{-1}$ would be needed. As can be seen all the amplifiers fall short of achieving this.

One further observation from this table is that the bandwidth is not important. The EL2008 had a much larger power bandwidth than the others but produced poor results. Clearly the input voltage, as would be expected, has the largest effect on the

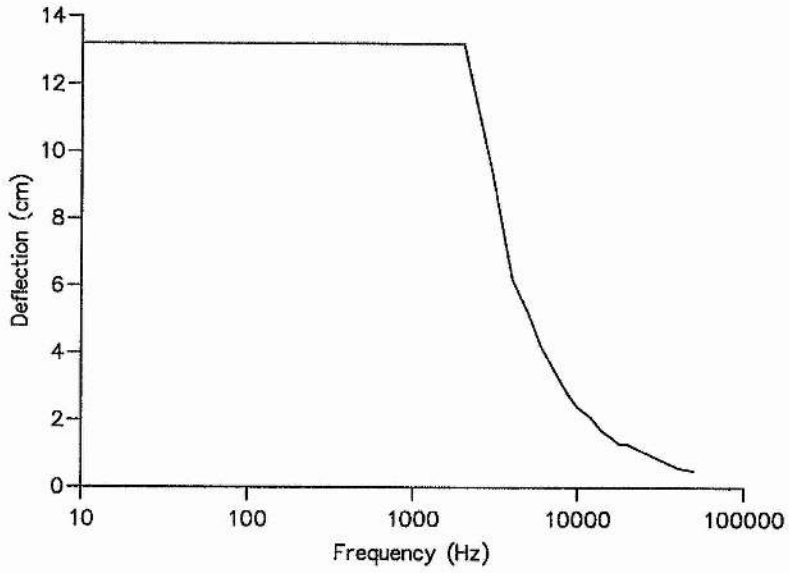


Figure 5.21: Graph of input frequency against beam deflection.

deflection.

Figure 5.21 shows the peak deflection of the beam plotted against input frequency for a driver stage using the LM12 amplifier [95]. Above 50KHz oscillations occurred and the drivers became unstable. The tailing off of the deflection is clearly the major problem that must be overcome in supplying high frequencies within a square wave. The LH0101 [96] is a higher bandwidth version of the LM12, this device could provide the added frequency response but, as it is now obsolete, having been superseded by the LM12, it was not available for test.

There are a number of ways to improve the high frequency response of these coils. The most obvious, although beyond the development facilities available at St. Andrews, is to redesign the yoke coils to optimize their bandwidth response. This would be unsuitable as a retrofit option in a commercial product but would solve the problem in the long term for specialized units. An alternative form of driving would be to use a secondary coil. This would contain as little as one turn, and would be wound in the same direction as the primary. By passing a high current through the secondary coil this coil could 'prime' the main coil. The secondary would be removed when the main coil current approached the correct level. This should accelerate the deflection and still

allow the normal coils to be left in place. This idea is based on work done by Rodime to enable fast settling of hard disk drive heads and thus reduce seek time.

5.3.6 Colour systems

In order to extend the application of video scrambling to a colour system it would be necessary to analyse the requirements of the three dot colour system used in most colour displays. In older systems three separate guns were used to create the three beams. This is easier to manufacture but is very bulky and has not proved adequate for modern display devices where many colours and higher resolution have made precision gun alignment essential and has led to the use of one gun and three beam gates. The extension of this invention to cover the older type of driving is, however, quite easy.

The three colour planes representing the red, green and blue images that are combined to give the final colour image are stored in computer RAM as three separate images. The three guns can be driven separately and the data removed from the three planes in the same way as has been specified for just one. If, however, the random sequence circuitry uses three separate sequences it would be possible to ensure that the guns were not necessarily at the same vertical position at any one time. This would not only allow colour displaying, but would also hinder the detection of the vertical positioning signals by electronic surveillance.

It is theoretically possible, in a single gun system, to detect both the intensity and the vertical signals. The colour system has three intensity signals and three vertical signals.

If detection was attempted at a distance greater than that needed to resolve the guns (the inverse square law applying to the gun emissions), the intensity received would be seen as the sum of the intensities from the three guns. As the displayed data does not have any correlation between the three guns (the red gun could be at the bottom of the screen and the blue at the top for instance), the three intensity signals cannot be separated. The vertical signals would be similarly overlayed, the single received waveform would have amplitude equal to the sum of the three amplitudes being displayed and no correlation exists that would enable the three values to be extracted.

Unfortunately the modern gun arrangement is to use three guns, but with common

vertical and horizontal deflection coils. The beams are much easier to align with each other, the system is cheaper and lower powers are needed. The mapping from one gun to three would now have to use the same vertical position for each beam. With modern systems allowing the use of up to 100Hz refresh rates it would be possible to scramble the beams sequentially, or even 2 at a time. This would reduce the quality of the display, but would render it almost perfectly secure.

5.4 Generating Random Sequences

5.4.1 Random and pseudo-random sequence generators

In order to implement the full patent specification it is necessary to generate a constantly changing scan order which in turn requires the generation of constantly changing random sequences. It is necessary to insure that all the numbers from 1 to N (where N is the total number of horizontal lines to be displayed) are generated, once only, but in an order that cannot be predicted. There are a number of well known circuits to do some of this, the simplest being the maximal-length shift register sequence generators.

Such generators produces all the numbers from 0 to $2^n - 1$ but in a fixed order. They work by starting at an arbitrary number then performing logical AND, OR and SHIFT operations on the bit pattern that represents this number. The result is the next number in the sequence. It doesn't matter where the start point is as the sequence is monotonic and the next number is defined by the current number being worked on, eventually the generator returns to the original start point, having generated all the numbers in the sequence.

This type of generator can easily be implemented in hardware. Figure 5.22 shows a simple implementation of a 50 bit sequence generator. Circuits of this type can easily be made into ASIC designs to make them more secure. The major disadvantage, however, is that the circuit can only produce a few different sequences. The example in figure 5.22 will repeat when all of the 2^{50} numbers have been generated. In reality, clocked at 10MHz, this will take 3·6 years.

In search of an alternative method, it became clear that a random number source would be easier to generate, and that algorithms that could generate random sequences from a random number or pseudo random number should be looked at.

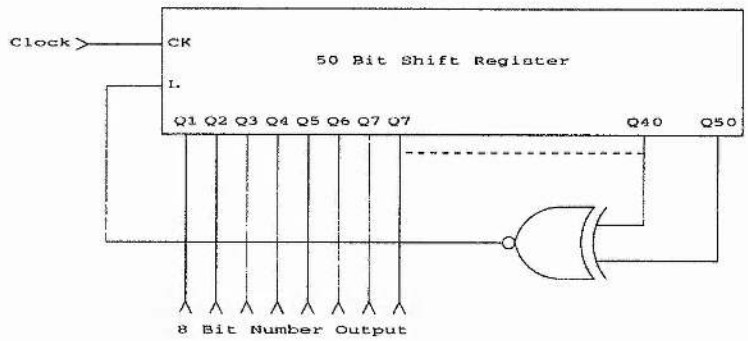


Figure 5.22: A 50 bit fixed sequence generator.

One simple pseudo random source is to use the sequence generator given in figure 5.22. If, say, an 8 bit random number is required, then any 8 of the available 50 bits could be used to make up an 8 bit word. With a repeating time of 3·6 years, this appears to be a good solution, and is in fact similar to most computer based pseudo random number generators. Alternatively, if an analog random source is preferable then just one bit could be fed out into a low pass filter circuit. By clocking the shift register at 10MHz, and filtering the output with a low pass filter set to cutoff at 1KHz, the output would contain uniformly random frequency components between DC and 1KHz.

The problem with both of these techniques, from a security aspect, is that once 50 bits of the register are known, it is possible to extrapolate forward and know what all the 50 internal bits are. This makes it possible to build a circuit that could synchronize its own shift register to the one being used to scramble the data. This also applies to the analog version though clearly a more complex decoder is necessary.

A more detailed look into the generation of genuine, uniform, random numbers will be covered in chapter 6.

5.4.2 Software solutions

Assuming a good, pseudo random number generator, a software algorithm can be developed that will produce a random sequence from a random number source.

Figure 5.23 is a program element illustrating how such a sequence is produced. In this case it is set to produce all numbers between 1 and 255.

```
(*-------------------------------------------------------------*)
(*----- Program element that produces a random sequence of -----*)
(*----- the numbers between 1 and max_value inclusive      -----*)
(*-------------------------------------------------------------*)

(*----- First create a flag for each possible value -----*)
for i:=1 to max_value do flags[i]:=0;
(*----- Next loop to output all max_value results -----*)
for i=max_value downto 1 do
    begin
    (*----- choose a random number between 1 and i -----*)
    num:=1+trunc(random(i));
    (*----- look for num flags that are still 0 -----*)
    count:=0;
    while num>0 do
        begin
        count:=count+1;
        if flags[count]=0 then num:=num-1;
        end;
    (*----- count is the result, set its flag & print it -----*)
    flags[count]:=1;
    write(count, ' ');
    end;
```

Figure 5.23: Program to produce random sequences.

The algorithm works by choosing a random number between 1 and the number of numbers still to be produced. For the first choice, this will be between 1 and 255. Once the first number has been chosen the new range will be between 1 and 254 and so on.

The next stage, once this random number has been chosen, is to set a counter to look at a flag for the number 1. If this number has already been output the flag will have been set. This will cause the counter to be moved on to the number 2. If, however, the flag has not been set, the chosen random number will be decremented by one, and then the counter will be incremented to now point to the flag for the number 2. Once the original random number has been decremented to 0, the current counter value will be output and then the flag that it is pointing to will be set. The next random number is now chosen, with the range reduced by one, and the process is repeated. Eventually all the flags will be set, indicating all the numbers have been output once and once only.

5.4.3 A proposed hardware solution

This system has the required ability to generate a random sequence, with the added advantages that it can operate for any number of elements in a sequence, and can be easily adapted to hardware implementation.

The software loops can be replaced with hardware counters and the software array can be implemented with a simple memory device. By adding a comparator, a clocked circuit can be built that will produce the required random sequence. Figure 5.24 shows a proposed implementation. It is designed to take an input between 0 and 255 inclusive, and produce a sequence ranging from 1 to 255. The multiplication required to generate a random number between 1 and M, from a random number that is always 8 bits, is accomplished by a ROM mapping. The ROM maps the input 8 bit number and the counter values (both fed into the address inputs) to produce an output according to the following equation.

$$\text{value} = 1 + \left(\frac{\text{random number}}{255} \times \text{counter} \right).$$

The principles are identical to those of the above program extract and for this reason it was felt unnecessary to develop this hardware any further. Once unscrambling hardware has been developed that extracts the data from the logical row, rather than use software scrambling, it might be interesting to research this idea further.

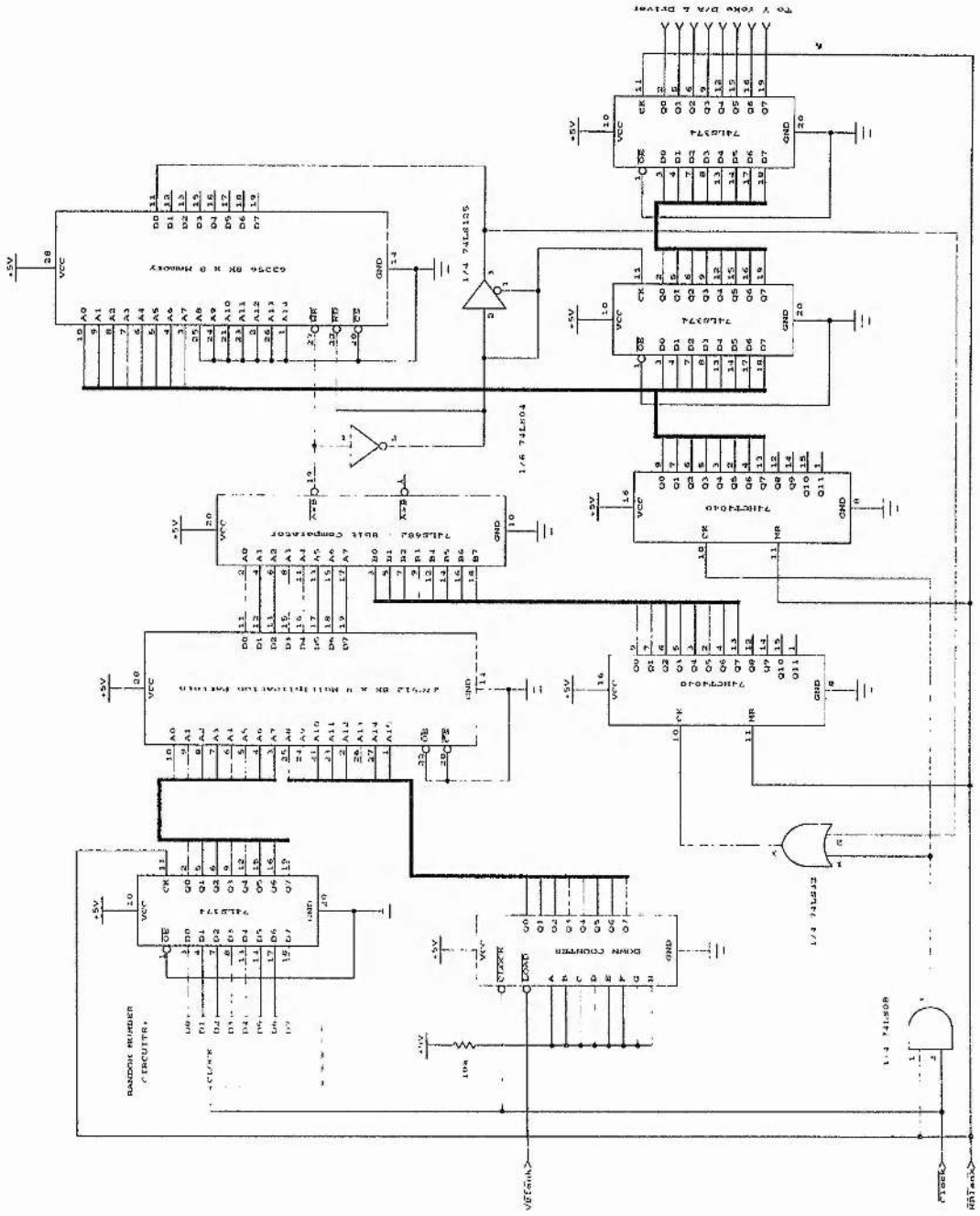


Figure 5.24: Proposed random sequence generator.

Chapter 6

Generating Uniformly Distributed Digital Numbers

6.1 Introduction

6.1.1 Objectives

There are many applications for random numbers including data protection, scrambling and cryptography [97]. In most cases a simple pseudo random number [98] will suffice but, particularly in the fields of security, a genuine random number generator is to be preferred.

There are a number of ways of generating such numbers and one of the most common is the use of a biased analog noise source [99, 100] to generate a random bit stream and then extracting words from this. Another commonly used method is to use a detuned FM receiver to generate high bandwidth white noise and extracting data from this. Systems such as these are good but suffer from the fact that an external source can interfere with them, in a known way. Though difficult, it is possible to predict their output when external conditions are changed from the norm.

Because of these problems there is a need to look at alternative methods of generating genuinely random numbers. It was noted that the generation of uniform numbers tended to rely on generating uniform bits and then building words from these.

A method is proposed here where a word is generated automatically from a random analog noise source. A conversion method is used to convert a Normally distributed voltage source into Uniformly distributed digital numbers. This provides an interesting offshoot into the theoretical analysis of the distribution and its underlying statistics.

This section will cover the theory used to produce these Uniform results. Later on analysis of the hardware needed to verify this theory will be undertaken and finally a detailed analysis of the distribution and the finite voltage cut-offs will be covered.

First however, it is necessary to look at the statistical conversion method used to get from Normally distributed samples to samples with a Uniform distribution.

6.1.2 Box-Müller and Normal distribution theory

A well known piece of statistical theory allows for the conversion of observations from a Uniform, $\mathcal{U}(x_{min}, x_{max})$ distribution to observations from a Normal, $\mathcal{N}(m, \sigma^2)$ by the use of the Box-Müller method [101, 102].

If r_1 and r_2 are independent observations from $\mathcal{U}(0,1)$ then it is possible to define

$$x_1 = \cos(2\pi r_2)\sqrt{-2\ln r_1},$$

and

$$x_2 = \sin(2\pi r_2)\sqrt{-2\ln r_1},$$

where x_1 and x_2 have an $\mathcal{N}(0,1)$ distribution.

By looking at the sum of x_1^2 and x_2^2 it can be seen that by simple manipulation of these expressions a reading, r_1, distributed as $\mathcal{U}(0,1)$, can be derived as

$$r_1 = e^{-\left(\frac{x_1^2+x_2^2}{2}\right)}. \tag{6.1}$$

In other words, with two independent observations from a Normal distribution it is possible to use equation 6.1 to get a single observation from a Uniform distribution.

The equation of such a Normal distribution is given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \tag{6.2}$$

where m is the mean and σ is the standard deviation of the distribution.

The following work is based around these equations and their application to discrete distributions where the full theoretical requirements of the Box-Müller conversions are not met.

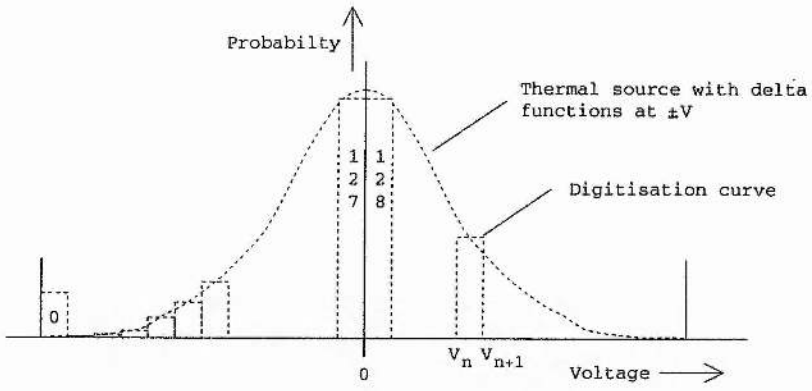


Figure 6.1: Digitizing a real world voltage source.

6.2 Calculation Of The Theoretical Parameters

6.2.1 Cut-off effects and scaling

The traditional Normal, or Gaussian, curve has the disadvantage in the electronics world of having limits at $\pm\infty$. Because the output of any noise generator must be constrained by the amplifier supply voltages ($\pm V$), a 'cut-off' form of the Normal curve must be generated when the voltage noise is sampled.

If a thermal noise source is used, the output voltage distribution will have this cut-off Normal probability function. This will have the same form as a true Normal function except at the limits $\pm V$. At the upper limit, V, the remainder of the probability function between this cut-off point and $+\infty$ will appear as a delta function at V and similarly for the $-V$ cut-off point.

If this voltage source is now digitized, i.e. it is split up into, say, 256 separate, equal spaced voltage regions, then the resultant probability density function will have these delta function added into the normal values for 255 and 0 respectively. This effect is shown in figure 6.1

This is not a standard distribution though it has some of the properties of the Truncated Normal [103], the Winsorized Normal [104] and the Censored Normal [105] where either out-lying points are ignored, or treated as uncertain. These distributions are covered in the literature but their analysis tends to concentrate on the effects they have on the honesty condition rather than on how the use of the new distribution is

affected by their differences from the Normal.

It is convenient to express the cut-off point in terms of the number of standard deviations from the mean rather than as a voltage limit. An observation x from an 8 bit digitization will now be defined as

$$x = \text{trunc}\left[\frac{V_{in}}{c\sigma_V} \times 128 + 128\right]$$

where c is the cut-off point measured in standard deviations and σ_V is the standard deviation of the voltage signal.

The effect of the TRUNC function is to subtract 0·5 from the distribution parameter mean, hence the approximation for this distribution is $\mathcal{N}(127{\cdot}5,(\frac{128}{c})^2)$.

In order to scale an observation from a Normal distribution, $\mathcal{N}(m,\sigma^2)$ to one from a distribution $\mathcal{N}(0,1)$ the readings must have the distribution mean, m, subtracted from them and then the result must be divided by the standard deviation, σ. This will now allow the use of equation 6.1.

By substituting

$$x_1' = \frac{(x_1 - m)}{\sigma},$$

$$x_2' = \frac{(x_2 - m)}{\sigma},$$

into equation 6.1, the result is an observation from the distorted uniform distribution $\mathcal{U}'(0,1)$.

This can be scaled to $\mathcal{U}(0,n)$ by simply multiplying the observations by n, and can be digitized by truncating the result. Equation 6.1 has now become

$$r = \text{trunc}\left[n \times e^{-\frac{(x_1'^2 + x_2'^2)}{2}}\right]. \tag{6.3}$$

It should be noted at this point that the probability of getting an observation $r = 1$ from a distribution $\mathcal{N}(0,1)$ is zero (as there are an infinite number of points along the line between 0 and 1). This means that the above conversion, when truncated, will produce all the integers between 0 and $(n-1)$ inclusive, but not n itself. This in turn means that in order to get an 8 bit result, the factor n must be 256 and not 255 as would be expected.

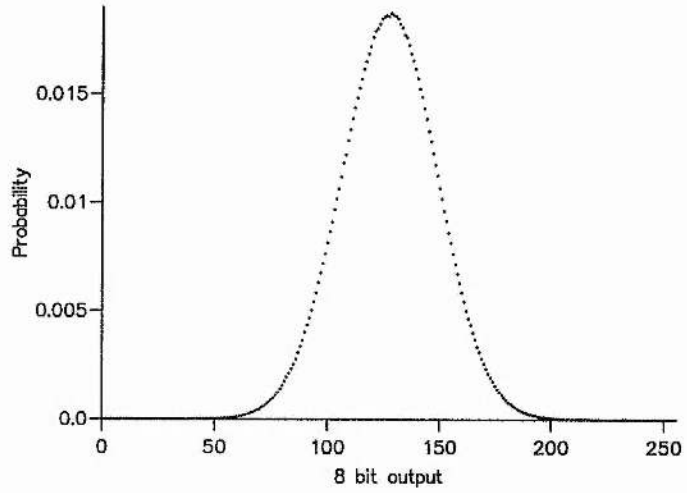


Figure 6.2: Graph of the pseudo-Normal distribution results.

6.2.2 Simulations

Having produced the above conversion theory, the next stage was to study the effect of the cut-off parameter to see how it could be optimized for this application. In order to achieve this a set of simulations were created that would take pseudo random, Normally distributed, numbers (from a simple Pascal function call) and map these to a Uniform distribution via the reversed Box-Müller conversion.

The voltage source was simulated by generating a random, real number from a shift and divide algorithm using 8 bit uniform random integers. In this way an almost continuous set of uniform observations from $U(0,1)$, were generated and these were then passed through the original Box-Müller conversion. This resulted in observations from an $\mathcal{N}(0,1)$ distribution. The correctness of this distribution was verified graphically and the results are shown in figure 6.2.

Readings from this distribution were then taken and processed via equation 6.3. This was done a large number of times for a fixed value of c. Other runs were then done using different c values.

In this way a visual representation could be analysed by eye and the most uniform distribution selected. The best value of c was found to be approximately 3·4 but the error on this was large and small variations in c caused a significant variation in the

'evenness' of the distribution. It was also necessary to look at a very large number of simulated points, over 1000 in most cases, before a true impression of the distribution could be seen. For the detailed analysis between 3·3, 3·4 and 3·5 this had to be extended to over 50000 points before a truly stable distribution was seen (taking 2·5 hours to calculate on a '286 PC).

It was noted that the quality of this output was very dependent on the number of units used to digitize the original, Normally distributed, voltage source. In all cases the results were scaled from the $U(0,1)$ into a $U(0,256)$ distribution and then truncated to simulate the generation of 8 bit digital numbers.

6.2.3 Evaluation of uniformity

One problem facing any attempt to optimize these equations using a computer based process is that the property of 'uniformity' for a distribution is very hard to define. Within statistical analysis there are a large number of methods of specifying such concepts, but only two are of any practical use in this situation, namely the maximum deviation from the mean and the sum of all the deviations squared.

The maximum deviation is of particular use when the distribution is close to the required format. All values will be close to the correct value (mean) and hence the worst case is a good measure of the quality of the distribution. However, if many of the points are correct but one or two are a long way out, then maximum deviation is not such a good measure and the sum of all the differences squared will be more relevant as it reduces the significance of individual values in relation to the whole.

For the initial work the sum of the differences squared was used to evaluate distributions and then, as the precise details became known the maximum deviation could be seen to be a better measure and was used for most of the more complex analysis covered in later sections.

6.2.4 Calculating a distribution analytically

Figure 6.1 shows a Normal distribution curve. The area between V_n and V_{n+1} represents the probability of an observation being between these two values provided the total area under the graph is equal to 1. As there is no explicit integral for the Normal density function, equation 6.2, then this area must be obtained by a very close approximating

equation.

The integral is known as the cumulative density function, Φ, and is defined as

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}\, dx = \int_{-\infty}^{x} P(x)\, dx.$$

Given this, and provided an approximation equation for $\Phi(x)$ can be found, then it is possible to find the probability of an observation being between V_n and V_{n+1} by

$$\mathrm{P}(V_n < X < V_{n+1}) = \int_{V_n}^{V_{n+1}} P(x)\, dx = \Phi(V_{n+1}) - \Phi(V_n). \tag{6.4}$$

If V_n and V_{n+1} are chosen so that they are the limits between which the A/D converter will give a reading of n, then, provided the Normal 'tails' are added into the probabilities of 0 and 255, this area corresponds to the probability that the input voltage will result in a digitized value of n being returned. The probability of getting a reading of n is now defined as

$$P(n) = \begin{cases} \Phi(n_{max}) - \Phi(n_{min}) & 0 > n < 254 \\ \Phi(n_{max}) & n = 0 \\ 1 - \Phi(n_{min}) & n = 255. \end{cases}$$

This is now in a form that can be analysed computationally and will be used later to investigate possible optimization algorithms. First however, approximations to the cumulative distribution function have be investigated.

6.2.5 Finding the cumulative density function, Φ

The limiting factor in calculations involving the cumulative distribution function, Φ, is the accuracy of the function used to approximate the Normal distribution function. Hastings' approximation [106] is the best known equation for this purpose. It states that

$$P(x) = 1 - Z(x)(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5) + \epsilon(x),$$

where

$$|\epsilon(x)| < 7{\cdot}5 \times 10^{-8}$$

and

$$t = \frac{1}{1 + px}$$

and the constants are

$$\begin{aligned} b_1 &= 0{\cdot}319381530, \\ b_2 &= -0{\cdot}356563782, \\ b_3 &= 1{\cdot}781477937, \\ b_4 &= -1{\cdot}821255978, \\ b_5 &= 1{\cdot}330274429, \\ p &= 0{\cdot}2316419. \end{aligned}$$

$Z(x)$ is the equation describing the $\mathcal{N}(0,1)$ distribution,

$$Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

This has the smallest error, $\epsilon(x)$, of all the approximations that were found for this calculation but, as it requires the value of $Z(x)$ (and hence the values for the constants e and π) to be known to at least 10 digits, it is not the most suitable for computation in all cases. A better, 'stand alone', function [107] in these instances is defined as

$$P(x) = 1 - \frac{1}{2}\left(1 + d_1 x + d_2 x^2 + d_3 x^3 + d_4 x^4 + d_5 x^5 + d_6 x^6\right)^{-16} + \epsilon(x),$$

where

$$|\epsilon(x)| < 1{\cdot}5 \times 10^{-7}.$$

These new constants, d_1 to d_6 are

$$\begin{aligned} d_1 &= 0{\cdot}0498673470, \\ d_2 &= 0{\cdot}0211410061, \\ d_3 &= 0{\cdot}0032776263, \\ d_4 &= 0{\cdot}0000380036, \\ d_5 &= 0{\cdot}0000488906, \\ d_6 &= 0{\cdot}0000053830. \end{aligned}$$

Having written procedures to test both methods for accuracy it was found that they behaved identically provided both were implemented in extended arithmetic in Sun Pascal. Figure 6.3 shows the form of $\Phi(x)$.

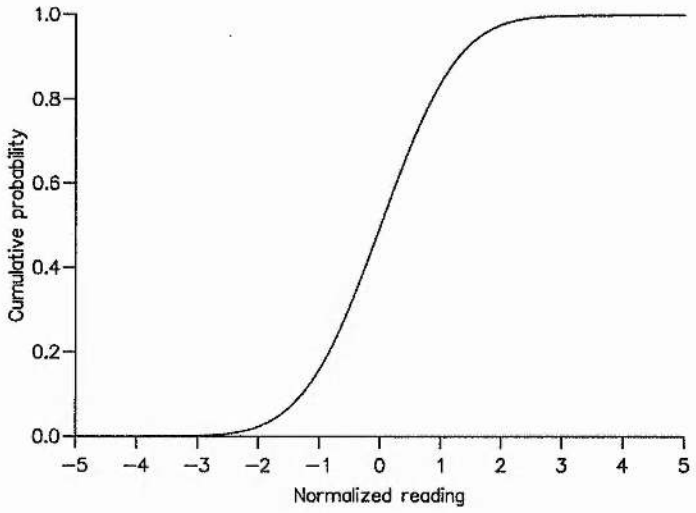


Figure 6.3: The Cumulative Distribution Function, Φ.

It was noted at this stage that there are a number of more complex approximations to this function where a set of 4 or even 6 equations are used to approximate different parts of the cumulative distribution. The Fortran NAG library contains an implementation of Hastings' algorithm given above but with a few extensions [108] to give better approximation values at the extreme points in the curve. It was decided for reasons of speed of execution that this should be used instead of the hand encoded versions though tests remain inconclusive as to which was the most accurate.

6.2.6 Analysing the effect of changing the cut-off point

With these approximations to $\Phi(x)$ it was possible to write a program that calculated the probability of obtaining any combination of inputs to equation 6.1.

By using two nested loops it is possible to enumerate these combinations and hence, using equation 6.4, calculate the probability of them occurring.

If these two input values are also passed through the Normal to Uniform conversion these probabilities will contribute to the probability of getting a particular results, n, from the conversion equation. If all combinations are enumerated and the probabilities of those combinations that produce an output value of n are summed, then the resulting total is the probability of getting n as an output value from the conversion equation. When calculated for all values of n (0 to 255) the probability density map has been

obtained.

As a check to the program, the honesty condition was invoked. This states that the sum of all the probabilities in both input and output distributions must be equal to 1. Having verified that this was so, the output distributions were compared to the simulation results found earlier. The correlation was very close and from this it was deduced that the routines were functioning correctly.

Having developed a method of calculating the distributions it was now possible to look at the effects of different cut-off values of the distribution, $\pm c$. Figure 6.4 shows how these changing values effect the distribution.[†]

Clearly at low values of c the conversion equation cannot produce small values of output. This, as will be seen in section 6.4, can be used to find the lower limit for the optimum value of the cut-off parameter.

As the value of c increases the quality of the distribution gets better up to a point somewhere between $c = 3$ and $c = 4$. After this the maximum deviations get worse until, at around $c = 5$ drop-outs appear where the equation cannot produce certain values.

Finally, at around $c = 16$, the conversion maps all inputs to small numbers and the larger values are not obtained. This provides an upper limit on the optimum value of c.

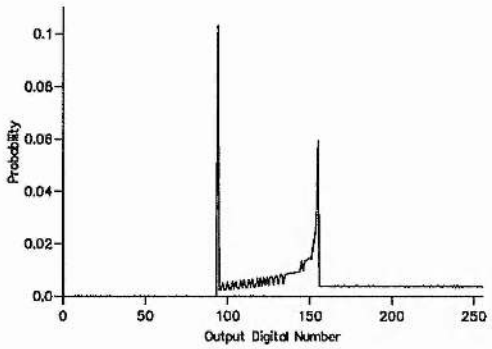
This technique of finding the distribution can be applied to look at larger numbers of bits used in the digitization of the noise voltage source, while still truncating the output of the conversion to 8 bits. Figures 6.5 and 6.6 show the distributions at $c = 3{\cdot}0$ for 9 bit and 16 bit digitizations respectively. As can be seen the increase in resolution causes a very significant improvement to the uniformity of the resultant distribution.

6.2.7 Finding c_0 from maximum deviations

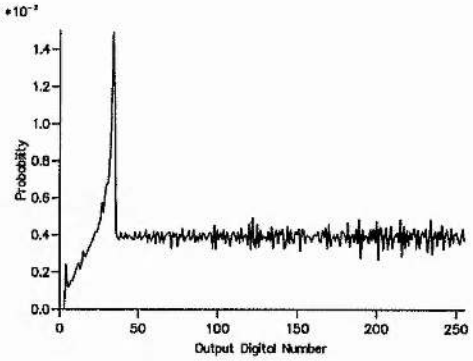
The is clearly some optimum value, c_0 for c. By computing all possible values of c, from 0·01 to 10, in steps of 0·01, it is possible to plot a graph of maximum deviation against c. This is shown in figure 6.7.

From visual analysis of this graph it can be seen that on the macro level, a value of c between 2·5 and 4·0 will produce the lowest deviations and hence the best conver-

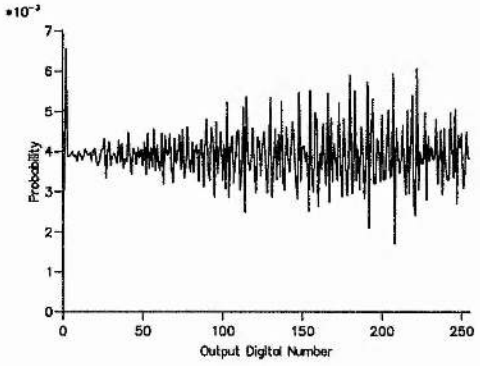
[†]Note the different axes that have been used for these illustrations



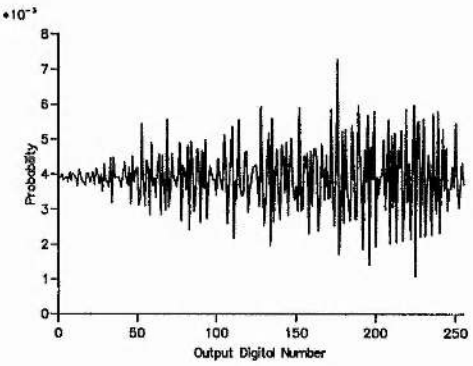
(a) c=1.0



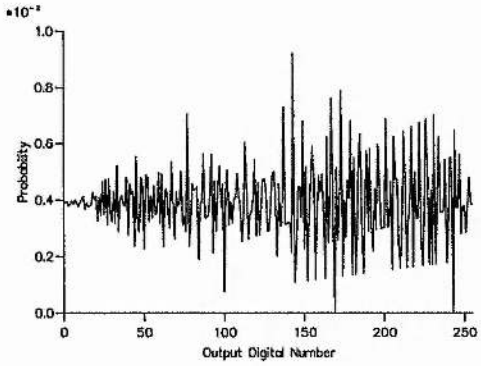
(b) c=2.0



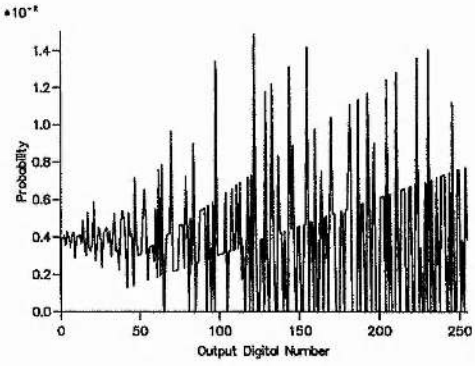
(c) c=3.0



(d) c=4.0



(e) c=5.0



(f) c=10.0

Figure 6.4: A selection of distributions for varying cut-off values.

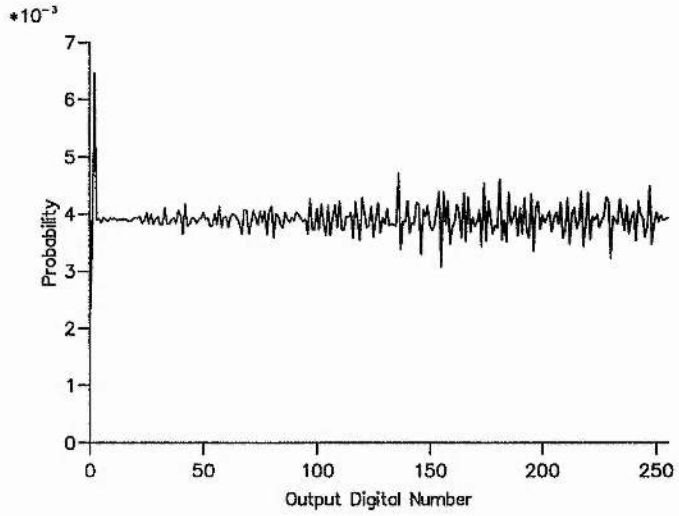


Figure 6.5: Distribution for 9 bits, $c = 3$.

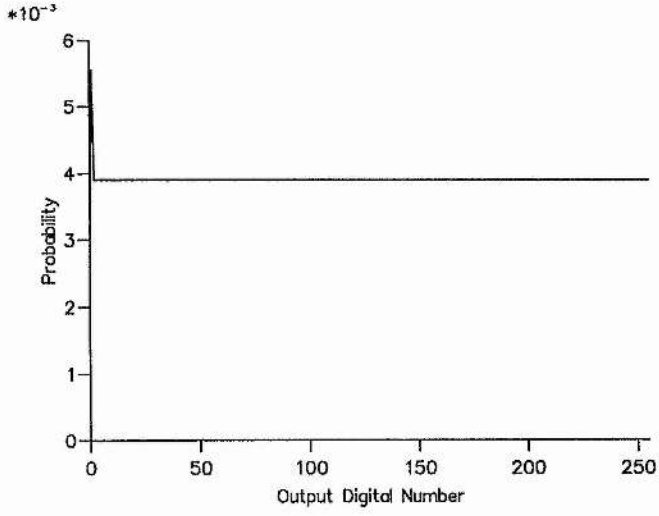


Figure 6.6: Distribution for 16 bits, $c = 3$.

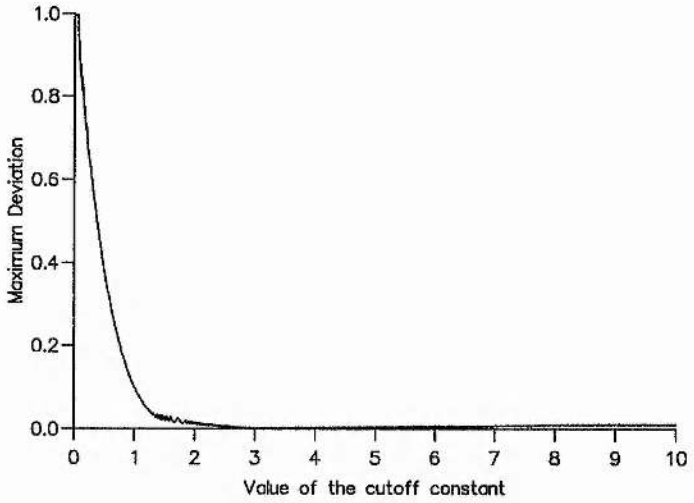


Figure 6.7: Graph of maximum deviation against c.

sion. As the gain settings for a simple hardware amplifier are only approximate this is more than accurate enough to specify the required value for experimental hardware verification.

6.3 Hardware Implementations

6.3.1 Noise generation

There are very few (if any) truly random events that are measurable, hence the more useful sources are those that cannot be determined without affecting the result, i.e. though not truly random, they cannot be predicted. One such source is a thermally produced voltage signal in a PN junction. Prediction of such a signal would require knowledge of the thermal states of every atom in the junction, which, even if it were possible to compute this prediction, could never be determined without affecting these states and hence the whole source is not deterministic.

There are some other systems that can provide pseudo-random output. These are mostly chaotic in nature [109] but have, as yet, no prediction method. However, in the analysis of such circuits their function is still deterministic relying on massive complexity to make them unpredictable.

Though there are many high frequency diodes that can be used to produce large

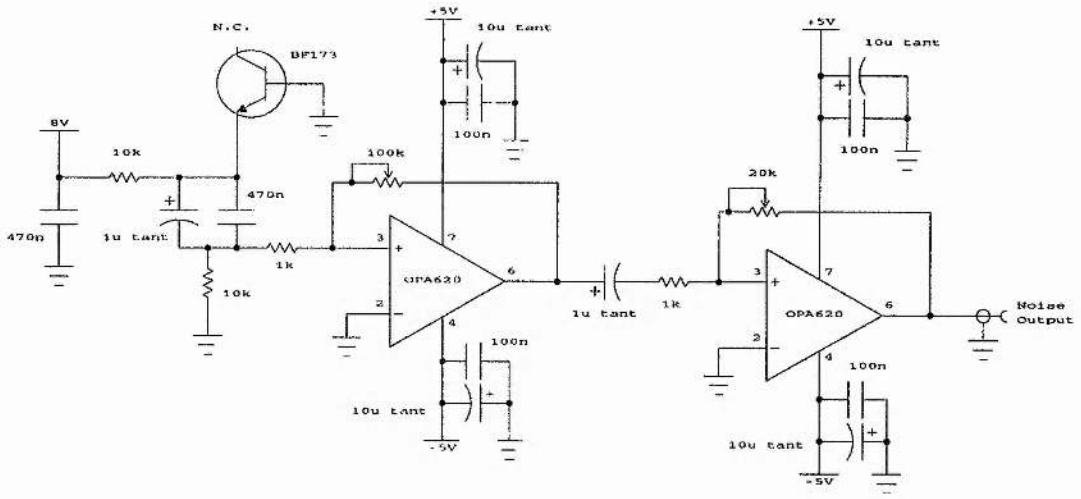


Figure 6.8: Circuit diagram of the wideband noise generator.

amounts of thermal noise, there are some that are particularly useful for this application. One such device is the BF173 transistor which, if reverse biased and left with a floating emitter, will produce noise at the collector-base junction which, amplified by the gain of the transistor, appears as a much larger signal at the emitter output.

The output of such a transistor based noise stage has a voltage distribution with standard deviation of 30μV and a noise bandwidth of $\sim$1GHz. In order to keep the bandwidth as high as possible, this had to amplified in small stages utilizing the higher small gain bandwidth of solid state amplifiers. The OPA620 was chosen as it had proved stable at many gain settings and has few pickup problems. It also has a wide full signal bandwidth.

Figure 6.8 shows the schematic of a circuit that implements this noise source.

A need for very good supply filtering was found, as the amplifiers became unstable at the least opportunity. By laying the circuit out on a long thin circuit board, the feedback coupling was kept below the point required to start oscillation and a final voltage shifter was then added off board to generate the correct voltages for input into an A/D converter. The total voltage gain was around $\times$20000 but the final stage was only a $\times$20 allowing an adjustable signal to be fed out to the converter and, coupled with the use of a multi-turn potentiometer, allowed accurate setting of the cut-off constant for the noise source.

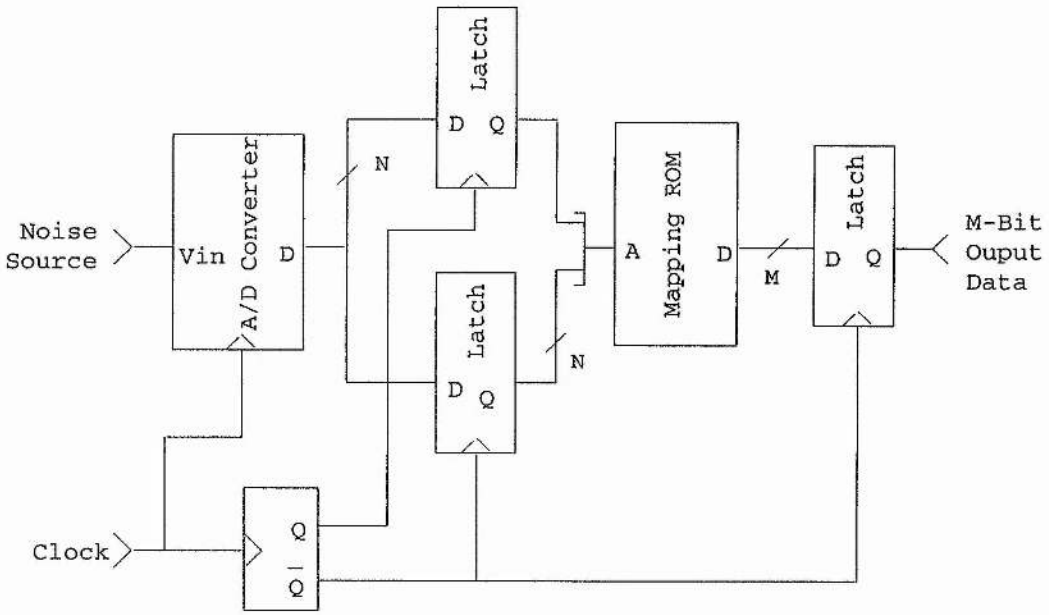


Figure 6.9: Block diagram of the sampling hardware.

In order to protect this circuit from outside interference, such as the processor clock of the PC, it was mounted into a UHF module that provided shielded inputs for the power supplies and also provided BNC connection for the noise signals. These also served to protect the sampling circuitry from radiated noise interference.

Having built this unit it was tested on a Hewlett Packard wideband spectrum analyser. The voltage output was seen to be proportional to the logarithm of the frequency, enabling accurate determination of the 3dB point of the circuit at $1{\cdot}7 \pm 0{\cdot}2$MHz. It was also noted that the signal did not reach the background until approximately 15MHz. The conclusion was that it would provide valid noise with frequency components of up to 5MHz. As this was large compared to the 0·5MHz requirements of the A/D converters being considered, it was valid to treat it as a source with infinite bandwidth as far as the input sampling circuitry was concerned.

6.3.2 8 Bit hardware

In order to obtain samples of this signal and then map them to Uniformly distributed observations, a sampling and mapping circuit had to be assembled. Figure 6.9 shows the block diagram of this circuit. The design is based around supplying the address

inputs of a normal ROM with two digitized noise signals. An input clock is divided into two phases, one latching a digitized signal to the top half of the address lines and then the other phase latching a digitized signal to the lower lines half a clock period later. This enables the ROM to be used as a mapping from two 8 bit inputs (the addresses) to one 8 bit output (the data lines). Once both data values have been latched, the corresponding output of the ROM is then also latched to give a valid continuous sequence of Uniform random numbers at the latch output.

Initially an 8 bit digitizer was used, the now familiar ADC304. This meant that a 64k byte ROM was required as there are 16 input bits, leading to 2^{16} possible combinations. As it was hoped to produce 8 bit values, a normal byte organized ROM was ideal and a 27C512 was chosen.

Figure 6.10 shows the final circuit diagram of the hardware that reflects the block diagram given in figure 6.9. A latch on the PC interface card, described in chapter 1, was now attached directly to the A/D output lines, allowing a PASCAL program to read this latch and produce a histogram of frequency of occurrence against reading for this circuit. This output was then used to calibrate the unit and set the c parameter.

6.3.3 Setting the c amplification parameter

In order to set the cut-off constant, it was necessary to analyse the results and feed back corrections to the amplification stages until the correct setting was achieved. This was done by collecting the data on the PC and plotting the distribution after every 10000 observations (about 5 seconds of collecting). Such a distribution is shown in figure 6.11.

An interesting feature of this figure is that some points occur slightly more often than others, in particular the probabilities for 8, 16, 32 etc are unusually high. An investigation into the cause of this produced two possibilities. One is that, as the converter uses a two stage settling process, it is possible that it is not sampling the signal exactly at the start of the conversion process and hence the input is changing before the first 3 output bits have settled to their correct values. The other possibility is that the converter has slightly different voltage 'windows' for each conversion number. This would be possible as the ADC304 is only specified to $\pm 0{\cdot}5$ units and hence the accuracy of each bit could be causing the problem.

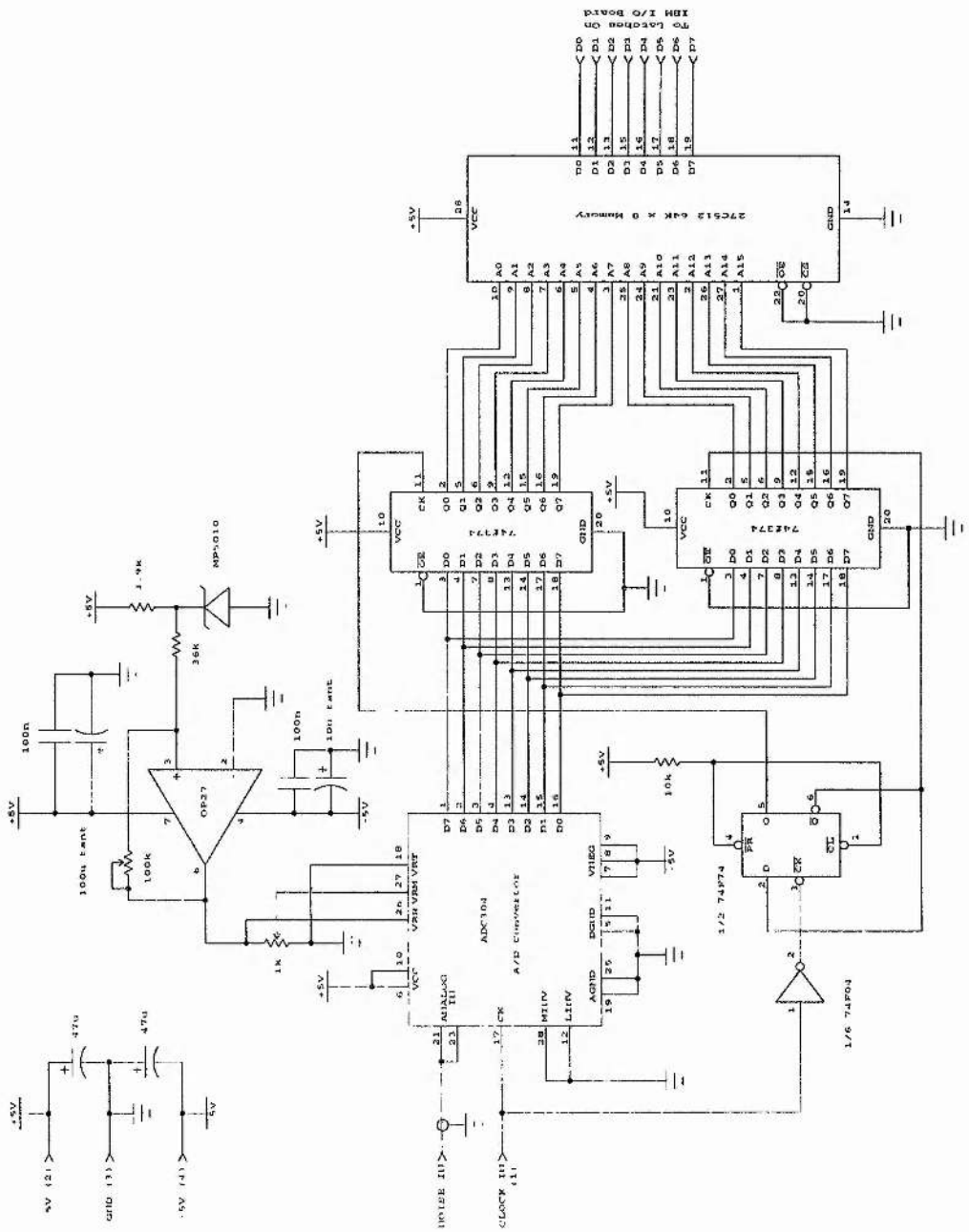


Figure 6.10: Full circuit diagram for the 8 bit mapping.

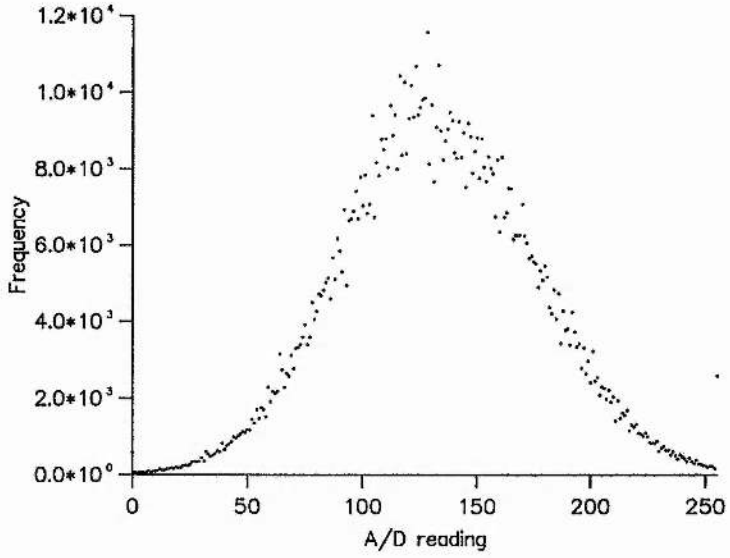


Figure 6.11: Example distribution from the converter data lines.

Fortunately both of these problems can be affected by external influences and improved results were obtained by reducing the input signal bandwidth. This was sufficient to prove the principals of the circuit.

Once these problems had been minimized the midpoint of the distribution was set by balancing the heights of the integrated 'tails'. This was found to be the best indication of where the midpoint was located. When both the 'tails' are the same height the distribution is exactly centered on the screen and hence correctly balanced with a mean of 127·5 as dictated by the theory.

Once the mean had been correctly set the peak height was measured, in millimeters, from the zero axis on the screen. The height of the distribution at $x = 1$ (i.e. the first reading from the distribution that is not part of the 'tails') was also measured as this corresponds to the height of the distribution at the cut-off point. These two points correspond to $f(\mu)$ and $f(x_c)$ respectively.

Because the cut-off point is defined in terms of the number of standard deviations it can be shown that

$$c = \frac{x_c - \mu}{\sigma}. \tag{6.5}$$

It can also be seen that the measured distances on the screen, because they have

been scaled, do not satisfy the honesty condition for a distribution. Because of this a scaling constant was added into equation 6.2 to get

$$f'(x) = \frac{1}{S\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

This can now be rearranged to get

$$-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 = \ln\left(f'(x).S\sigma\sqrt{2\pi}\right), \tag{6.6}$$

and since

$$f'(\mu) = \frac{1}{S\sigma\sqrt{2\pi}}, \tag{6.7}$$

equations 6.5, 6.6 and 6.7 can be combined to get

$$c = \sqrt{2\ln f'(\mu) - 2\ln f'(x_c)}.$$

As an example, the average of 6 such measurements of $f'(\mu)$ and $f'(x_c)$ were 54·2 and 5·3 respectively, and hence c =2·16. Clearly the gain was set slightly too high in this case as the cut-off point is too close to the mean.

Alternatively the reverse of this process gives

$$e^{\frac{c^2}{2}} = \frac{f(\mu)}{f(x_c)}.$$

If c is set to 3, the ratio $\frac{f(\mu)}{f(x_c)}$ is equal to 90, i.e. the ratio of the peak reading to the minimum reading is 90:1.

Once the correct amplification and offset (mean adjustment) had been achieved, a much larger sample set, 10^6 measurements, was taken to verify the distribution.

6.3.4 Generating a mapping ROM

In order to build the mapping hardware a program had to be written to generate the complete 2^{16} values needed for the 27C512 EPROM (and later the 27C020 256k×8 bit component [110]). This program enumerated all the possible input values using two nested loops ($X1$ and $X2$), running from 0 to 255. This in turn was used to generate the addresses by the equation $A = 2^8 X1 + X2$.

The equations given in the last section lead to the mapping from $X1$ and $X2$ to the output result, r. These were used to create the data values for the ROM cells.

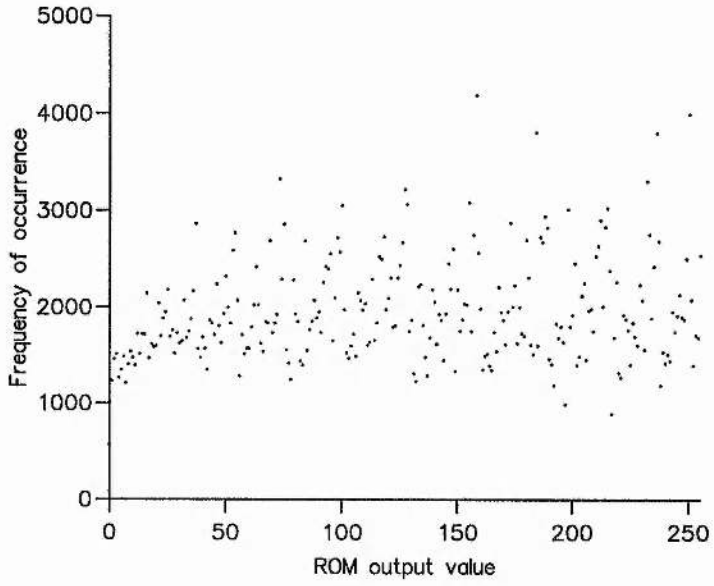


Figure 6.12: Output from the Normal to Uniform mapping hardware.

As the addresses were generated sequentially, the program was able to directly create Motorola 'S' record output that could be fed directly into the EPROM programmer.

By viewing this data as a contour map it can be seen that the middle point (127,127) has a very high probability of occurring but, as readings are taken further from this midpoint they reduce in probability. As the mapping is symmetrical in both X_1 and X_2, any points on a circle around this midpoint will have the same probability. This is the inverse of the normal distribution curve, as would be expected.

6.3.5 Results from the 8 bit hardware

Once the parameters had been set for the ROM mapping, the distribution data could be downloaded into the PROM programmer and then into the ROM. By moving the 8 bit sampler connections on the PC interface card to the output latch of the circuit the Uniform data could be sampled and uploaded into the PC in a similar way to the collection of Normal data used earlier. Figure 6.12 shows the output for 10^6 samples. As can be seen it is very close to the expected level of 'uniformity' predicted in earlier sections.

Device	Resolution	Conversion Method	Price	Speed
AD573JN	10 bit	Successive approximation	£27·90	30μs
ZN502E	10 bit	Successive approximation	£13·00	20μs
ADC1061CIN	10 bit	Modified Successive approx.	£19·90	1·8μs
MP7685KD	10 bit	Flash Conversion	£45·68	2μs

Table 6.1: High resolution converters..

6.3.6 9 Bit hardware

The next stage was to look at generating a more Uniform output by using 9 bit inputs to the mapping ROM. This is about the limit of the hardware as the two 9 bit inputs require a ROM with 18 address lines. This corresponds to 256k bytes and was the maximum EPROM size available. To make use of the bandwidth available in the noise source the circuit also requires a fast 9 bit converter. Table 6.1 shows a selection of suitable converters. In the end it was decided to use the MP7685KD supplied by RS Components [111]. This had the required conversion rate and also had the advantage of using a full flash conversion process. It was also hoped that this would overcome some of the problems encountered with the earlier versions of the circuit using the ADC304.

An extra flip-flop was added to latch the extra two address bits in parallel with the 8 bit latches used earlier. Another latch was also needed to enable all 9 bits to be read into the PC and stored.

Unfortunately the same problem of some values having unusually high probabilities reappeared for this version of the hardware. Again small changes in the surrounding circuitry helped to minimize the effect. A more detailed analysis of the distribution around the problem points showed that these peaks were set in slight troughs. This meant that for some reason values close to these peaks were sometimes converted as if they were on the peak. In fact the peaks always occurred on the borders of powers of two - for some reason the last few bits were going to zero in some instances.

Changes in the constant voltage reference supplies, in particular adding a high current driver, helped reduce the problem and showed that another possible cause of the problem was noise from this reference voltage. Large decoupling capacitors on this input made significant improvements and enabled the rest of the circuit to function

without any degradation in performance.

This 9 bit conversion circuitry, including the high current drivers for the reference source, is shown in figure 6.13. Figure 6.14 shows an example output distribution from this hardware. If compared with that of figure 6.12 it can be seen that the improvement, predicted in the theory, has been shown in practice.

6.4 Finding An Exact Value For c_0

6.4.1 A more detailed examination of the distribution

When the equations for the distribution are analysed, more details of the structure of this distribution can be seen.

The conversion into a Uniform distribution maps pairs of input values to single output values in a many-to-one relationship. This is because the Normal distribution is symmetrical and any Uniform output value, generated by (x_1, x_2) will also be generated by (x_2, x_1) and also by taking identical points mirrored about the mean of the Normals. This means that there are at least 4 pairs of inputs per possible output and in some cases there are more than this (though always in multiples of 2).

During the mapping process the input pairs form an output value on the real number axis. These values can be represented as delta-like functions with an area equal to their corresponding joint probability, $p(r) = p(x_1) \times p(x_2)$. Figure 6.15 shows a section of the real axis illustrating how these delta functions occur.

From this diagram the next stage of the conversion can be seen. The truncation and probability summing operation is illustrated in this diagram by summing the areas of all the delta function between N and $N + 1$ where N is an integer.

With low values of the cut-off constant, all these delta functions have high positions on the real axis. They all start between 255 and 256 but as the cut-off value is increased these peaks begin to spread out down the axis, always keeping their same relative positions, but expanding downward. This expansion gradually moves peaks out of the 255 to 256 band and some now fall in the range 254 to 255. Eventually the peaks spread out into the 0 to 1 band and all region values from 0 to 255 will have some peaks in them. The optimum positioning of these peaks is dependent on the area of the delta functions and the number of such functions, in each region.

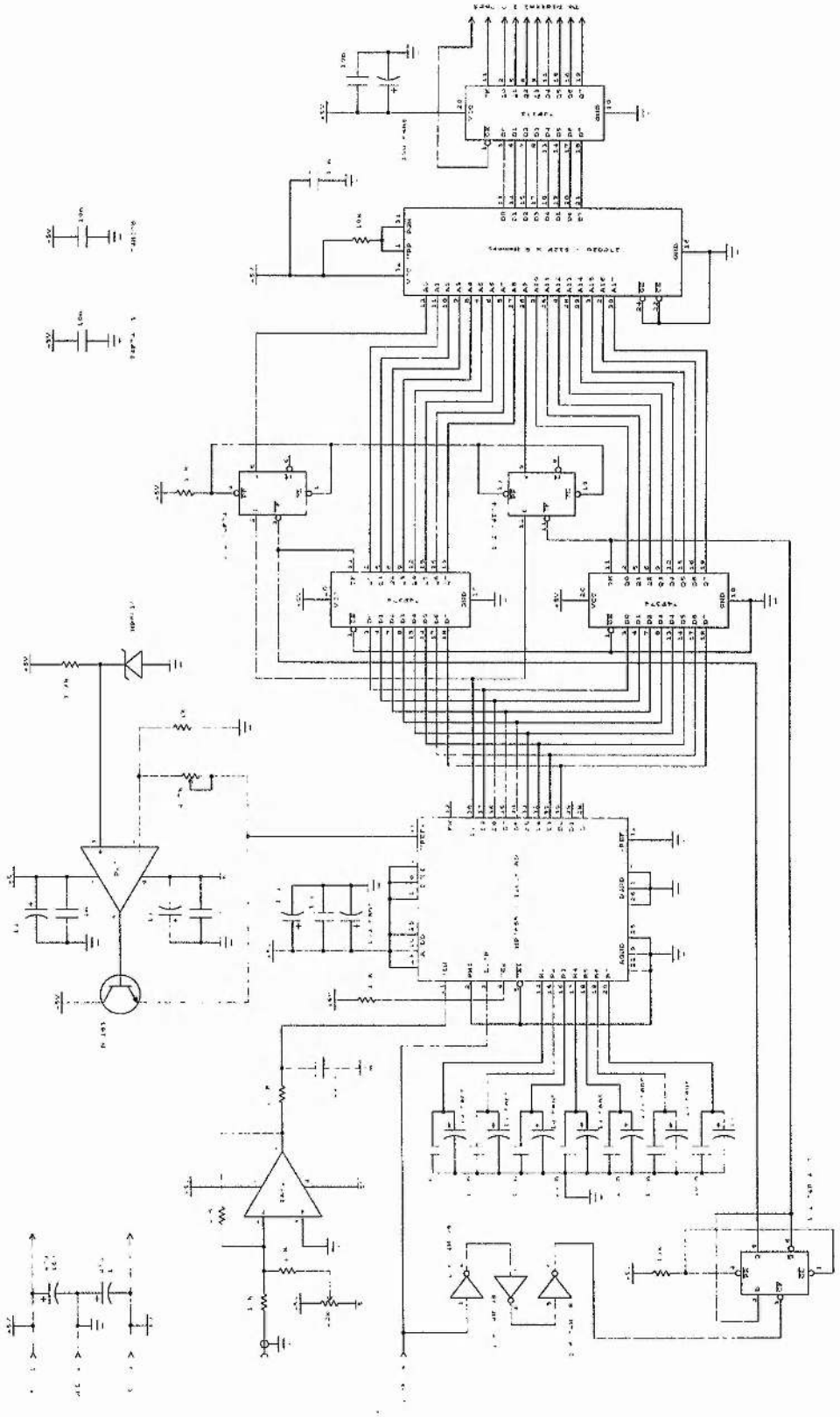


Figure 6.13: Full circuit diagram of the 9 bit mapping hardware.

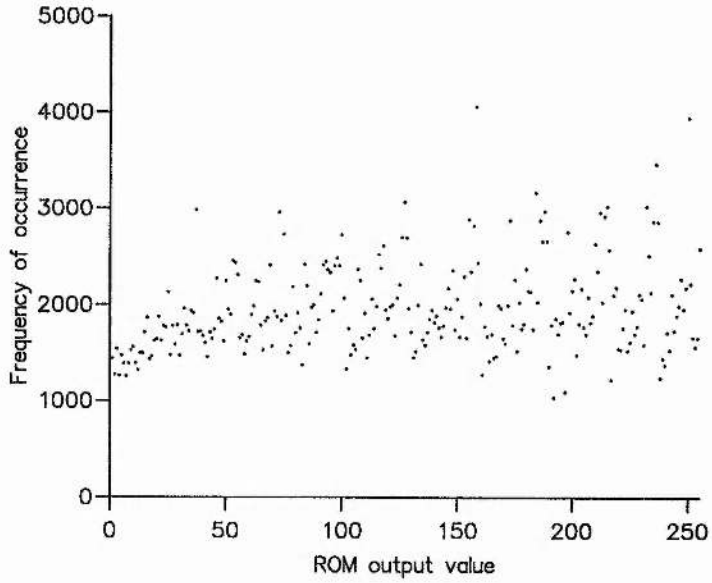


Figure 6.14: Results of the 9 bit Normal to Uniform conversion.

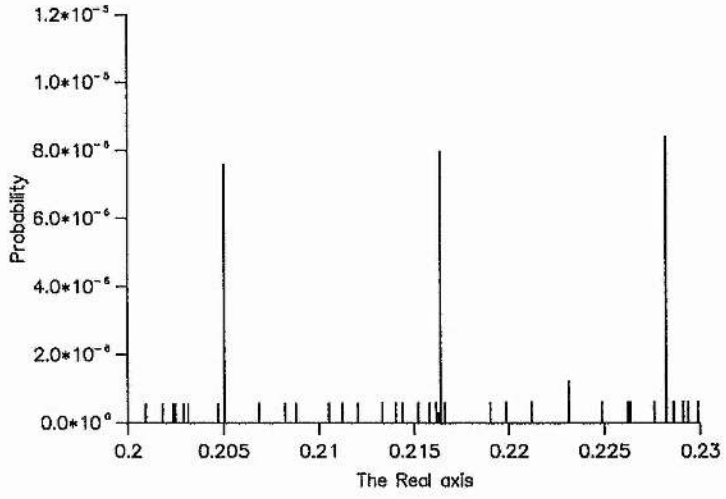


Figure 6.15: Part of the real axis showing the distribution of the delta functions

This analysis shows that there is a minimum value of the distribution. Figure 6.4c shows that it is possible to have a distribution where the maximum deviation from the ideal Uniform is less than $\frac{1}{256}$. This implies that there must be a peak in every section of the real axis as, if a group had none, its probability would be 0.

From equation 6.3, the minimum value of r, before truncation, is found when the scaling value x' is a maximum and this occurs when $x_1 = x_2 = 0$. By substituting this into equation 6.3 it can be shown that

$$\begin{aligned} r_{\min} &= 256e^{-\left(\frac{x-127\cdot5}{128}\right)^2 c^2} \\ &= 256e^{-0\cdot9922c^2} \end{aligned}$$

for an 8 bit result.

From this, since this must be less than 1 for all groupings 0 to 255 to have some peaks in them, it follows that

$$-0\cdot9922c_0^2 < \ln\left(\frac{1}{256}\right)$$

and hence

$$c_0 > 2\cdot36.$$

Going back to figure 6.15, as the cut-off is increased the tail end of the real axis becomes less populated and finally it will fall outside the band 255 to 256. At this point there is a similar definition for the maximum value of c_0.

For this scaled Normal, the result is

$$255 \leq 256e^{-\frac{c_0^2}{256^2}},$$

and hence

$$c_0 \leq 256\sqrt{\ln\left(\frac{256}{255}\right)},$$

resulting in the maximum bound of

$$c_0 \leq 16\cdot02.$$

6.4.2 Analysis of the distribution using integers

For any given c value the distribution of the delta functions can be found and hence the groupings for values between say, 0 and 1 can also be determined. Having derived

the groupings for $c = 2{\cdot}36$, the minimum value, the next possible grouping is when one, and only one, of the delta functions moves down the real axis, passing over an integer boundary and joining a lower group.

This value was found by looking at every delta function and calculating the corresponding c value that would just put it into the next lowest group. This meant that, after a possible new c value had been found for every delta function, the minimum such value could be determined and hence a new c value could be used and the new distribution calculated.

This is clearly a long and complex process so a few improvements were needed in order to reduce computation time. Firstly, to reduce problems of calculating so many points an integer solution was sought.

By expanding the calculations, the position on the real axis of the resultant, r, is found to be

$$r = 256 * \exp\left[-\frac{1}{2}\left(\left[\frac{(0{\cdot}5 + x_1) - 128}{128}c\right]^2 + \left[\frac{(0{\cdot}5 + x_2) - 128}{128}c\right]^2\right)\right].$$

This is because the integer values have to be scaled to be from an $\mathcal{N}(0,1)$ distribution and also because the value representing any given region is taken to be half way through that band. For example when the integer value is 127, the actual range is between 127 and 128 and hence the mean of 127·5 is used.

This equation can be rearranged to generate integer solutions and the best reduction is found to be

$$\frac{65536}{c^2}\left[\ln(r) - \ln(256)\right] = 510(x_1 + x_2) - 2x_1^2 - 2x_2^2 - 65025. \tag{6.8}$$

Using this, integer solutions for all the peak positions can be found (the right hand side of the equation) and only the mapped integer values (i.e. the left hand side for r=0, 1, 2 etc) need be found as real numbers.

It is now possible to look at only the nearest peak to the mapped integer value in each group rather than look at all possible peaks. This has to be carried out for all regions as some peaks move faster along the real line than others, though they never overtake each other.

In effect, now that integer solutions have been found, the peaks are fixed and it is the region boundaries that move.

Another improvement was to find both the nearest peak to a divider and also the second nearest. These two values were used to allow the dividing point to be set to half way between the two corresponding c values, ensuring the smallest effect of rounding errors. If the minimum value were used then the peak that is supposed to have moved into the next section may not actually do so when the calculation is done.

Because it is now possible to determine which peak crosses the divide first, provided a known grouping can be established at the start, then there is no need to actually work out the new locations. The heights (areas) of the peaks merely need to be recalculated for the new c value and the summing done by the known, rather than calculated, groupings.

In some cases the mathematical resolution was so poor, even using 32 bit arithmetic, that it was not possible for the analysis program to decide which peak moved into the next group first. One way to overcome this problem is to analyse all the possible combinations of moving one peak at a time. Only one of these combinations would be the correct one and the rest would give false values as the distribution would be momentarily wrong. Provided none of these combinations produces a minimum value for the deviation, this approach can be used to ensure that all possible distributions are examined.

This latter stage was not actually implemented as this program merely aimed to improve the known value and not deduce it precisely and the added processing time was not acceptable.

When run this program produced an answer of 3·11637552399035. This was found to be caused by the (89,123) peak passing from the 164 region to the 163 region. It also corresponds to a maximum deviation of $1{\cdot}7252946484629 \times 10^{-3}$.

6.4.3 Deducing c_0 from the minimum combination

By considering the distribution around this point it can be seen that the maximum deviation grows smaller as c reduces over this range. This means that the minimum value occurs at the previous point where this peak just crosses the divide line. This can be calculated analytically by rearranging equation 6.8 to get

$$c = \sqrt{-\ln\left(\frac{164}{256}\right)\frac{1}{2}\frac{256^2}{(x_1 - 127{\cdot}5)^2 + (x_2 - 127{\cdot}5)^2}},$$

which yields a c value of 3·116373099.

6.4.4 Further analysis of the distribution

The problem with this approach is that it does not take into account the fact that the maximum deviation changes with c value within a grouping as well as between groupings. By looking at the distribution the major effect on the minimum deviation is found to be due to peaks moving between groups but there is still this small change within a group, caused by the $\mathcal{N}(0,1)$ scaling.

As the honesty condition must be satisfied for any distribution, this change is not linear, some peak areas increase and some decrease with c. The overall effect is that a program must be written to look at both the changes within a grouping and the different groupings themselves and then find the maximum point for each of these possible distributions. This would be far too time consuming with available computing facilities, since the previous program to calculate c_0 for just the best grouping took over 11 days to run on a Sun Spark 2 system.

A program written in the Maple symbolic algebra package was able to find the minima by looking at the value changes. Unfortunately this is slow for one group (a few seconds) and could increase the total computing time to months if not years for a complete calculation

It is very unlikely that the results of this program would come up with any different answer from that already obtained. The effects on the distribution of changing c are small within a grouping while the difference between distributions when a peak changes region is much larger and hence the effect of altering c, when the groupings do not change, is not likely to be significant. However, until such a proof is obtained the value cannot be guaranteed as being correct.

6.4.5 One final improvement

Having calculated a possible value for the minimum deviation it is possible to reduce the original range by a small amount. For large values of c the largest peak is generated for $x_1 = 127$ and $x_2 = 127$. By symmetry this peak is in the same place as the peaks for (127,128), (128,128) and (128,127).

For smaller values of c the tail effects dominate and the largest peak is found to

be generated from $x_1 = 0, x_2 = 0$ and the four combinations of 0 and 255 (the other 'mirrored' tail effect).

Given the largest peak, it is known that this cannot be larger than the maximum deviation possible, as this peak alone would contribute too much to the distribution. The value calculated in the previous section gives one possible minimum deviation for the distribution and hence the height of the largest peak must be less that $\frac{1}{256}$ plus this value since the deviation, not the absolute value must be used.

Hence, for the first peak (127,127)

$$4 \times P^2(127) \leq \frac{1}{256} + 1{\cdot}72529464846290 \times 10^{-3},$$

and since $\Phi(0) = 0{\cdot}5$ it can be seen that

$$\Phi\left(\frac{-c}{128}\right) \geq -\frac{\sqrt{\frac{1}{256} + 1{\cdot}72529464846290 \times 10^{-3}}}{2} + 0{\cdot}5,$$

and for the other peak

$$\Phi\left(\frac{-127}{128}c\right) \leq -\frac{\sqrt{\frac{1}{256} + 1{\cdot}72529464846290 \times 10^{-3}}}{2}.$$

A program was written that found the inverse of the cumulative density function, Φ^{-1}. This was then used to find both a minimum and maximum value of c depending on which peak is used. This resulted in the range of c_0 being improved at the upper limit and combining this with earlier results the range for c_0 is

$$2{\cdot}36 \leq c_0 \leq 12{\cdot}06.$$

Clearly the c_0 parameter does not, for all practical purposes, need to be precisely defined. Any value approximating the above calculated optimum will ensure that mapping hardware will produce useable Uniform distributions from the cut-off Normal inputs present in a real world thermal noise source.

Chapter 7

Conclusions

7.1 Possible Future Work

There are still a large number of alternative scanning methods available for both the capture and display of images. The ideas of using area CCDs as one dimensional devices can be taken much further and the applications of unusual scanning displays have only just been touched upon.

In order to do further work into novel scanning methods a full electromagnetic display system would be essential. The preliminary work done in chapter 5 shows some of the difficulties that such work would have to overcome and also points at some ways of overcoming them. High bandwidth coils are easy to design given the required simulation software and the expertise and facilities to build and test coils. It is doubtful that modifications to the existing coils could achieve the required bandwidth for anything other than specific, purpose built, devices but it would be interesting to see the limits of modern multi-scan monitors where y-axis scanning of over 200Hz is already available.

The uses of protection systems such as the random scanning system described earlier are numerous and many people have expressed interest in such devices including the Law Society and the American Embassy. The development of a VGA system would be advantageous for the marketing of such a system and the prototyping of a colour system would be essential.

The hardware developed to look at laser pulses needs some fine tuning to get the correct coupling and frame shifting that would be needed in a commercial product. Possible improvements, including the use of ground plane PCBs and putting the supply

regulators onto the camera head itself, show that the design can be improved and the addition of automated software analysis would provide suitable driving for ease of use.

The X-ray one dimensional camera can also be improved by the use of commercial PCB design and would benefit from more work being done on the scanning and readout electronics, particularly with reference to the methods of recognizing a single X-ray event. It may be that a software algorithm could be developed to look at the resultant images or that some in-line circuitry could be used instead, possibly using a 3×3 grid to carry out signal processing.

There appears to be no reason why the replacement of the optical CCDs with X-ray sensitized devices should cause any problems.

The surface generation hardware should be easily adapted to work from a normal video signal, though this would limit the accuracy and applicability of the camera in the form it has been developed so far. Again the use of PCBs would enhance the quality and should also allow the device to run at its full crystal speed by reducing the crosstalk pickup on the output signal.

Finally the obvious improvement to the theoretical work is to prove that the deduced cut-off point is correct and to extend this proof to find such points for all numbers of quantization. It does not seem likely, due to the discontinuous nature of the conversion process, that an analytical solution can be found but there seems to be a good chance of finding a proof for any specific quantization level.

Such a proof may well be found by looking at the maximum possible change in the value of c between the limits for each possible grouping. This would allow a proof to be deduced if it can be shown that such limits confine the 'looked for' value to only one grouping. This would then lead to an analytical solution once this grouping is known.

To reduce the limits further it may be possible to scan through the possible groupings, looking at the total probability of peaks in the 0 to 1 group. This must always be less than the deviation used to deduce the upper limit in section 6.4.2. Such a program would reduce the limits still further but is still specific to the groupings for 8 bit values and is not as general as it could be.

7.2 Conclusions

All the projects here have proved the principles of their operation and have verified that the ideas can be made to work and that the results are as predicted.

The surface generation is a good example of turning computer programs into hardware in such a way that a vast increase in throughput is obtained. It also illustrates how much can be done by approximations, particularly as no calculations are needed for the final result whereas many real number operations were needed for the original rendering.

The scanning methods used for the security project allowed further development and helped draw attention to the problems faced by bodies such as the Law Society. The continued interest in this project may well lead to some commercial realization, with at least the verification of the idea clearly illustrated by this work and by subsequent testing done at Marconi Electronics.

Both the one dimensional cameras have proved that CCDs can be successfully used in one dimensional applications and that modification to the standard scanning methods can reveal very interesting properties of both CCDs and their associated output processing electronics. Both devices have obtained reasonable results and can easily be turned into commercially useful electronics.

The problems associated with dark current and digitization have been examined with some success and it is interesting to note the effects these have on the results, particularly since they are often ignored by people acquiring data without appreciating the errors inherent in their acquisition methods. It is hoped that some headway has been made into rectifying this.

The discovery of a constant defining the optimum cut-off point for digitized use of the Normal to Uniform conversion leaves open a number of interesting questions and proved a very challenging piece of research, with interesting results.

It is to be hoped that at least some of the right questions have been asked, and that some of them have, at least in part, been answered.

Appendix A

Charge Coupled Devices

A.1 Introduction

The CCD has largely taken over as the optically sensitive element in modern camera systems providing normal TV images to conventional displays using either PAL or NTSC encoding. There have been a large number of manufacturers interested in these devices, including Sony, RCA, EEV, Thomson-CSF, Schlumberger, Texas and Hamamatsu along with many smaller players in this large market.

Today there are hundreds of different devices on the market and they are almost all dedicated to the TV or video camera markets. They have enabled high street shops to sell very high quality cameras for less than it would have cost for just the CCD five years ago.

This appendix aims to give a brief introduction to CCD technology and the concepts behind driving their inputs and processing the resultant outputs.

There is now a wide range of applications for CCDs within the scientific community. Originally astronomy applications took the lead, using the rigidity and re-usability of the CCD to replace photographic plates. In particular, the RGO [112], the Kitt Peak observatory [113] and the Galileo Institute for astronomy [114] did some pioneering work on the cooling of CCDs to get long time exposures. The fixed arrangements of the cells is particularly useful to this application and allows correction of image plane curvature in telescopes with curved focal planes [115]. It was also within the astronomical community that the first work into thinning CCDs was done. This allows back illumination and hence reduces the photon loss through the front electrodes [116].

Large format CCDs with anything up to 1242 × 1152 pixels, such as the EEV

CCD05-30 series [117], have recently been under development for dental and small area X-ray work. Small pixel (10μm) devices [118] have been developed for high definition cameras and 'stackable' devices have been used to create very large area sensors.

A.2 CCD Structure

Both CCDs and their cousins the diode arrays [119] are based around silicon structures that have been sensitized to convert photons to electrons and then localize these electrons to store an image.

The structure of these devices has gone through many revisions since their invention in the late 1960s but all are based around the same basic principals. A typical device consisted of a silicon substrate with a set of parallel, equally spaced, vertical channels covering the whole surface. Onto this is laid a set of horizontal, parallel electrodes at right angles to the channels, the result being a chequered effect of electrodes and channels covering the silicon surface.

The channels act as potential barriers that prevent horizontal drift of electrons stored in the silicon. The horizontal electrodes provide charge collection fields which are also used to move the electrons vertically across the surface.

There are a number of approaches to charge collection in these devices. Most are based on the fact that after an optical or X-ray photon has passed into the silicon (through the electrodes) its energy is absorbed by the doped silicon depletion layer and electrons are produced.

These electrons will be attracted to any electrode that has a positive charge applied to it. Traditionally the three phase model has been used to describe the charge collection strategy. There are however many other possibilities and Texas devices in particular have been leading the field in the use of two phase clocks.

The three phase model uses the electrodes in groups of three. The middle one is kept at a high potential and the outside two are held near ground to form an electron 'bucket'. The next group of three will have a similar pattern of voltages on them. When an electron is formed in the depletion layer it will be attracted to one of the positive electrodes. As there are two low voltage electrodes between each positive one the electrons will be kept separate once they have migrated in the positive field.

If the voltage levels on the electrodes are now changed so that the bottom of the

three electrodes is positive and the top two are at a low potential, the electrons stored under the center electrode will be pulled down by the new field and will sit below the bottom electrode.

Again the two low potential electrodes will prevent any mixing of stored electrons as they will always have a strong field either holding them in place or pulling them in one direction. In the 'bucket' analogy the 'bucket' has been moved down.

This system allows electrons in the entire image to be moved down one pixel row, that is through 3 electrodes for each full three phase clock cycle.

The normal frame based CCD has two silicon areas. The top half is a 388×288 area segmented into pixels by the groups of three electrodes. This area is exposed to the outside and is where images are formed by an external lensing system in a camera.

Below this section is a similar area, the same size, but covered by an opaque metal shield. This prevents incident light from affecting the active silicon surface.

Electrons form in the top half of the sensor during the frame period (0·02 seconds for a normal PAL TV signal), creating an image of electrons in the silicon. The electrode shifting system can then be used to move the entire image, quickly, into the lower half of the device. This process is done as fast as possible during a vertical blank period to minimize the smearing caused by incident photons falling on the sensor during the frame shift process.

Once resident in the storage area the image is shifted one row at a time into a bottom row readout register. Here the electrodes are at right angles to the image movement (replacing the channels in the image areas) and allow the electrons to be shifted horizontally to a pixel output register. Here they are amplified by a FET transistor stage and then fed to user external circuitry, the nature of which will depend on the application.

This process of reading out one row and then shifting down the stored image to allow the next row to be output is done at the pixel rate for the display system. For a PAL TV this is approximately 8MHZ.

Once the entire image has been read out the cycle repeats with the top image again being moved to the storage area.

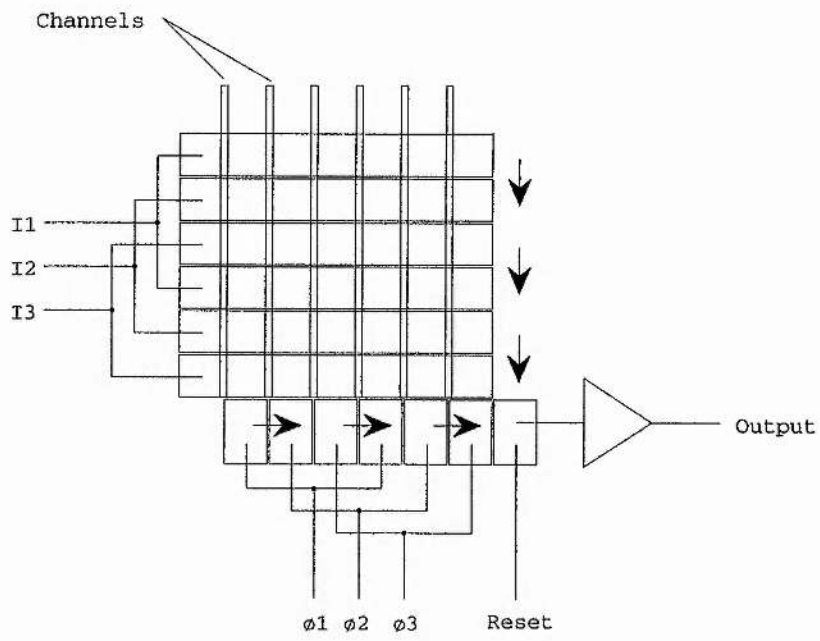


Figure A.1: The electrode patterns on a CCD surface.

A.3 Driving CCD Inputs

As with all other aspects of CCD technology, there are many possible ways of driving the voltage and shifter inputs with the astronomical applications leading the field [120, 121, 122]. In order to drive the electrodes, three sets of three inputs are provided. All the top electrodes in a group of three within the imaging area are connected to an output pin, all the middle electrodes to another and the bottom electrodes to a third. These outputs are then mirrored for the bottom (storage) section so that the image and storage sections can be moved independently.

Finally the row readout electrodes are brought out in a similar way with every third electrode being connected together. A simplified example of this structure is given in figure A.1.

To ensure efficient transfer of electrons the driving waveforms must be sloped so that the electrical fields overlap slightly between the electrodes. This ensures accurate electron transfer, leading to good charge transfer efficiency. These slopes can be created by small capacitive filters attached to a digital, square wave driver or for more precise readouts a full ramp generator can be employed.

For scientific purposes CCD readout waveforms are usually of much lower frequency allowing more complex and more precise forming of the pulse shapes. Normal camera systems however usually rely on careful setting of low pass filters. In some cases special techniques such as slight negative biasing of the electrodes can be used to maximize the transfer efficiency and charge collection ability of the electrodes particularly while the electrons are not moving.

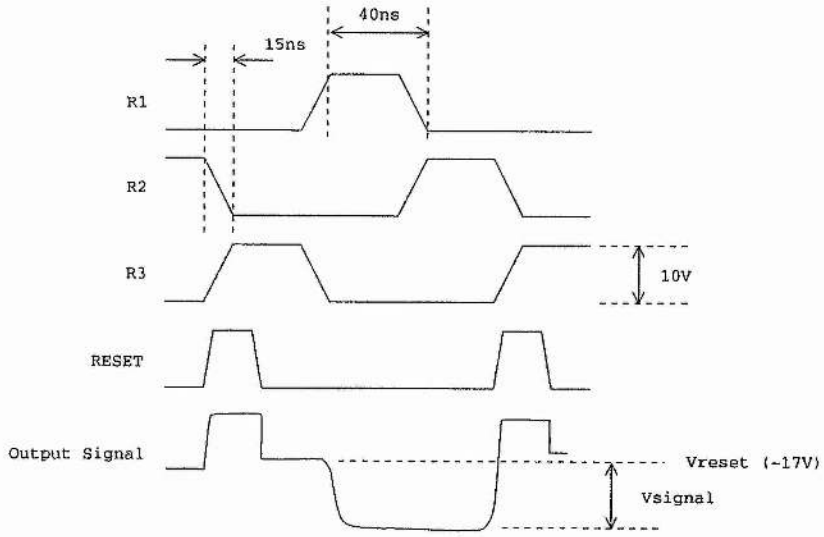
A.4 Output Processing Methods

There are a number of approaches to CCD output processing. Many companies are beginning to offer ASIC signal processors and some useful work has been done into producing very low noise pre-amplification stages [123, 124] for CCDs and reticons.

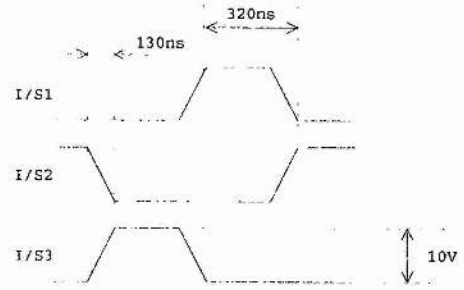
Conventional CCDs have two signal outputs. These are both fed from identical transistor charge amplifiers on the silicon and hence have very similar output properties. One of these amplifiers is fed with the charge from the output register whereas the other is tied low and has no input charge. This second output acts as a dummy voltage level that can be used to correct for output distortions in the other, live, signal.

When the charge has been transferred to the CCD output amplifiers it becomes the signal that is passed on to the outside processing electronics. This charge must be removed from the output amplifier before the next charge packet arrives. This is achieved by one extra waveform input that feeds in a positive signal to the charge input stage of the amplifier and removes the electrons that have been stored there. As this pulse is never perfect, this reset level changes slightly after each pixel is read out. This means that the output waveform is quite complex. During the reset pulse period the output is at its highest voltage level. Once the reset pulse is released, the output sinks to a quiescent state, with slight noise levels introduced by the imperfect charge removal of the reset pulse. The output remains at this level until another charge packet is transferred to the readout stage. This will appear as a negative voltage and will pull the output down. The required signal is the difference between this new low voltage and the final level of the quiescent state. An example output waveform is provided in figure A.2a along with the phases of the row readout waveforms and the column shift signals (figure A.2b).

The bias voltages to the base silicon and the electrodes are also quite complex,



(a) Row shifts



(b) Column shifts

Figure A.2: Typical CCD driving waveforms and output signals.

different levels being required for the gates, amplifiers, substrates and channels. One effect of this system is that it induces a large DC offset on the output drivers which can be as high as 17V. This, along with the reset noise, must now be removed by the user processing electronics before a valid signal can be obtained.

The advantage of the dummy output is now apparent. This output has both the 17V DC offset as well as the reset noise found on the real signal output. A good differential amplifier stage attached to the real and dummy signals can now subtract the dummy signal and produce just a voltage representing the optical signal levels.

One problem with some implementations of this is that the DC offset cannot be handled by the outside video amplifiers and is often removed by capacitive DC blocking filters, even though distortions will be incurred by such filtering. This is fine for family camera systems but not of much use to scientific quality instrumentation. This is a major problem with many supposedly 'high quality' camera systems.

One other way of removing the reset noise is the double correlated sampler. Here the reset signal level is sampled in its quiescent period and fed into an integrator for a short period. The actual signal, including the reset noise is then negatively integrated for the same period. The result of this is that the output signal from the integrator has had the noise level removed (canceled out by the positive and then negative integration) but since the signal is only present for one half of the integration, it is still available as an integrated output.

The basic noise removal systems are improved by the reduction of the local temperature. The thermal noise in the readout stages can be reduced to almost zero by cooling to liquid nitrogen temperatures and the best astronomical instruments used in this way reduce the total noise to just 3 or 4 electrons per pixel, almost achieving single photon counting [125]. Smaller temperature reductions can be achieved by using Peltier cooled CCDs that are now supplied by firms such as EEV [126] specifically for astronomical work.

CCDs are also sensitive to other parts of the spectrum [127]. They are particularly sensitive to infra-red radiation and, though sensitivity drops rapidly for ultra violet detection, special coatings [128] can be used to obtain good results for the higher optical frequencies. Sensitivity also improves for high energy radiation and peaks sharply for X-ray wavelengths. The silicon gives very high gains for the conversion of X-ray photons

into electrons, particularly in the soft X-ray regions and even to direct neutron radiation [129].

The problem with using such wavelengths, and in fact the direct illumination of CCDs with free electron beams [130], comes in the form of radiation damage to the CCD surface and in particular to the covering electrodes. The silicon is also not capable of trapping all the energy of the incident photons when it is optimized for optical sensitivity and this leads to the non-recording of some of the incident X-ray energy.

There are two techniques employed in CCD manufacture to improve their response to X-rays. Firstly the devices are thinned and irradiated from the back [131, 132], and secondly they are built from high resistivity silicon that traps more of the energy and is usually made as a thicker device to enable better localization of electrons. Both of these systems improve the application of CCDs to X-ray detection and are part of the reason that most of the significant advances into CCD uses today are in the areas of synchrotron radiation detection and medical imaging.

There are many more detailed works associated with CCDs. In particular manufacturers data sheets such as those produced by EEV [133] and RCA [134] provide very good insights to the field. There are also a number of review articles [135, 136, 137] and most modern electronics texts cover at least some aspects of CCD technology [138, 139].

References

[1] Leggatt D. P. *'The Evolution of television technology'*, Electronic Eng. 60(735), 14, 1988.

[2] Moffat B. *'Television engineering research in the BBC, Today and Tomorrow'*, SMPTE Journal-Society of motion picture and television engineers. 97(1), 17-24, 1988.

[3] *'Advanced Video Display Controller (AVDC) SCN2674'*, Mullard, Aug 1983.

[4] *'CCD02-06 Series Scientific Image Sensor'* EEV limited, (1), June 1990.

[5] *'The PNA7518 8-bit Multiplying DAC'*, 289-293, Aug 1987.

[6] *'LM733/LM733C Differential Video Amp'*, National Semiconductor, 9-54 - 9-57.

[7] *'CD54/74HC4040, CD54/74HCT4040 12-Stage Binary Counter'*, (1483) 375-379.

[8] *'DS0026/DS0056 5MHz Two Phase MOS Clock Drivers'*, National Semiconductor Corporation, 5-13 - 5-20.

[9] *'ESB365089AA Buffer Hybrid'*, EEV Company Ltd., March 1987.

[10] *'ESB365091AA'*, EEV Company Ltd., July 1987.

[11] *'SHM-360, SHM-361 Video Speed Sample-Holds'*, Datel Inc. 3-7 - 3-10.

[12] *'Quad Differential Line driver Quad TTL to ECL Translator'*, Signetics, P73.

[13] *'ADC-304 8-Bit, 20MHz LOW POWER, FLASH A/D'*, Datel Inc., 1-33 - 1-38.

[14] *'Ultrahigh-Frequency Operational Amplifier AD5539'*, Analog Devices, (9) Feb. 1987.

[15] *'HOS-050/HOS-050A/HOS-050C Fast Settling Video Operational Amplifiers'*, Analog Devices, 2-254 - 2-250.

[16] *'OPA620 Wideband Precision OPERATIONAL AMPLIFIER'*, Burr-Brown IC Data Book, (33), 2-166 - 2-169.

[17] Arndt U. W. *'The collection of single-crystal diffraction data with area detectors'*, Journal De Physicque. 47(8), c5.1-c5.6, Aug 1986.

[18] Thomas D. J. *'Calibrating an Area-Detector Diffractometer'*, Modern Diffractometry series 2(2) 1988.

[19] Arndt U. W. *'X-ray television area detectors'*, Nucl. Intrum. and Meth. 201, 13-20, 1982.

[20] Arndt U. W., In'T Veld G. A. *'Further developments of an X-ray television detector'*, Adv. Electronics and Electron Phys. 74, 285-296, 1988.

[21] Mathieson A. McL. *'Small-Single-Crystal Diffractometry with Monochromated Synchrotron Radiation - the Wavelength-Dispersion Minimum Condition for Bragg Reflection Profile Measurement'*, Acta. Cryst. A444, 239-243, 1988.

[22] Mathieson A. McL. *'Concerning Single Crystal Reflectivity Curves'*, Aust. J. Phys. 41, 393-402 1988.

[23] Arndt U. W. *'Measurement of the* 3D *distribution around a single-crystal Bragg reflection'*, Internal communication, July 1986.

[24] Germer R., Mayer-Llse W. *'X-ray TV camera at 4.5nm'*, Rev. Sci. Instrum. 57(3), 426-427, March 1986.

[25] Germer R. *'High resolution X-ray-TV-sensors'*, S.P.I.E., 491, 434-491, 1984.

[26] Naday I., Strauss M. G., et at. *'Detector with charge-coupled-device sensor for protein crystallography with synchrotron x rays'*, optical engineering 26(8), 788-794, Aug. 1987.

[27] Westbrook E. M., Deacon M. L. *'CCD-Based area detectors for protein crystallography'*, A.C.A., July 26 1988.

[28] Borso C. S. *'Optimization of monolithic solid state array detectors for the position encoding of small angle X-ray scattering from synchrotron sources'*, Nuc. Instr. and Methods, 201, 65-7, 1982.

[29] Galileo Electro Optics Corporation. *'Galileo Glass Capillary Arrays'*.

[30] Bigler E., Polack F., Lowenthal S. *'Scintillating Fiber Array As a X-ray Image Detector'*, 1989.

[31] Strass, M.G., Naday, I. et. al. *'CCD Sensors in Synchrotron X-ray Detectors'*, Nucl. Instr. and Method in Phys. Research, 578-577, 1988.

[32] Arndt U.W. *'X-ray Position-Sensitive Detectors'*, J. Appl. Cryst., 19, 145-163, 1986.

[33] Strauss, M.G., Naday, I. et. al. *'CCD Sensors in Synchrotron X-ray Detectors'*, Nucl. Instr. and Methods in Phys. Research, 266, 563-577, 1988.

[34] Lumb, D.H., Holland, A.D. *'X-ray imaging spectroscopy with EEV CCDs'*, S.P.I.E., 982, 116-122, 1988.

[35] Lumb, D.H., Chowanietz, E.G., Wells, A.A. *'X-ray Imaging with GEC/EEV CCDs'*, Conference paper.

[36] Kellog, E., Murray, S. et. al. *'The Photicon'*, S.P.I.E., 290, 28-33, 1981.

[37] Gruner, S.M. *'CCD and Vidicon X-ray detectors: Theory and Practice (invited)'*, Rev. Sci. Instrum., 60(7), 1545-1551, July 1989.

[38] Koppel L. N. *'Soft X-ray response of a charge-coupled image sensor'*, Rev. Sci. Instrum., 48(6), 669-672, June 1977.

[39] Lumb D. H., Hopkinson G. R., Wells A. A. *'X-ray imaging and spectroscopy with CCDs'*, Adv. in Electronics and electron Phys. 64(b), 497-507, 1985.

[40] Stern R. A., Liewer K., Janesick J. R. *'Evaluation of a virtual phase charge coupled device as an imaging X-ray spectrometer'*, Rev. Sci. Instrum. 54(2) 196-205, 1983.

[41] *'Pixel Detector Workshop'*, Synch. Rad. News, 3(3), 6-8, 1990.

[42] Catura R. C., Smithson R. C. *'Single photon X-ray detection with a CCD image sensor'*, Rev. Sci. Instrum. 50(2), 219-220, 1979.

[43] *'Radiation damage effects in EEV CCDs'*, technical note 1991

[44] Magorrian B. G., Allinson N. M.*'Soft X-ray Damage In CCD Detectors'*, Nuc. Instr. and Methods in Phys. Research A273, 599-604, 1988.

[45] Peckerar M. C., McCann D. H., Leepo Yu. *'X-ray imaging with a charge coupled device fabricated on a high resistivity silcon substrate'*, Appl. Phys. Lett. 39(1), 55-57, July 1981.

[46] Walton D., Stern R. A. et al. *'Deep-depletion CCDs for X-ray astronomy'*, S.P.I.E. 501, 306-316, 1984.

[47] Griffiths R. E. *'The Evaluation of Silicon CCDs for imaging X-ray Spectroscopy in the Range 1 to 8 keV'*, Adv. in Electronics and Electron Phys, 64(b), 483-496, 1985.

[48] Lumb D. H., Hopkinson G. R., Wells A. A. *'Performance of CCDs For X-ray imaging and spectroscopy'*, Nuc. Instrum. and Meth. in Phys. Research. 221, 150-158, 1984.

[49] EG&G Reticon. *'S Series Solid State Line Scanners, 128, 256, 512, and 1024 Elements'*, 1-37 - 1-46, Sept. 1987.

[50] Zutavern F. J., Schnatterly S. E. et al. *'A position-sensitive photon detector for the UV or X-ray range'*, Nucl. Instrum. and Meth. 172, 351-355, 1980.

[51] Gamble R. C., Baldeschwieler J. D. *'Linear position-sensitive X-ray detector incorporating a self scanning photodiode array'*, Rev. Sci. Instrum. 50(11), 1416-1420, Nov 1979.

[52] Borso C. S., Danyluk S. S. *'Application of a directly exposed self-scanning photodiode array as a linear position sensitive detector in a small-angle X-ray scattering instrument'*, Rev. Sci. Instrum. 51(12), 1669-1675, Dec 1980.

[53] Widom, J., Feng, H.-P. *'High performance X-ray area detector suitable for small-angle scattering, crystallographic, and kinetic studies'*, Rev. Sci. Instrum., 60(10), 3231-3238, Oct 1989.

[54] Chu, B., Wu, D.Q.. *'Evaluation of a linear array detector for synchrotron small-angle X-ray scattering measurements'*, Rev. Sci. Instrum. 60(10), 3224-3230, Oct 1989.

[55] Luppino, G.A., Ceglio, N.M. et. at. *'Imaging and nondispersive spectroscopy of soft x rays using a laboratory X-ray charge-coupled-device system'*, Optic. Eng., 26(10), 1048-1054, Oct 1987.

[56] Jucha A., Bonin D., et al.. *'Photodiode array for position-sensitive detection using high X-ray flux provided by synchrotron radiation'*, Nucl. Instrum. Meth. in Phys. Research. 226, 40-44, 1984.

[57] Integrated Device Technology Inc. *'High Speed 1k x 9 Dual-Port Static RAM with Busy - Preliminary, IDT7010S/L. IDT70104S/L'.*

[58] *'American institute of physics handbook'*, Section 8.4.

[59] Fork R. L., Greene B. I., Shank C. V.Appl. Phys. Lett. 38, 671, 1981.

[60] Finch A., Sleat W. E., Sibbett W. *'Subpicosecond synchroscan operation of a photochron IV streak camera'*, Rev. Sci. Instrum. 60(5), 839-844, May 1989.

[61] Sibbett W., Sleat W., Taylor J. R. *'Application of synchronously scanning streak cameras to picosecond time resolved luminescence measurements'*, Picosecond Chemsitry and Biology. 197, 1983.

[62] Sibbett W., Sleat W. E., Krause W. *'A picosecond streak camera For spaceborne laser ranging'*, Proc. 5th Int. Workshop on Laser Ranging and Instrumentation. 1, 136-164, Sept 1984.

[63] Baggs M. R., Eagles R.T. et al. *'Design & Applications of Photochron Streak Cameras'*, Conf. on Photoelectronic Imaging. 48, 253, 1985.

[64] Diels J. C., Fontaine J. J., et al. Appl. Optics. 24, 1270, 1985.

[65] Eagles R. T., Sibbett W., Sleat W. E., Walker D. R. *'Multiple-frame UV/X-ray picosecond framing camera'*, Adv. in Electronics and Electron Phys. 74, 209-217, 1989.

[66] Boyle W. S., Smith G. E. *'Charge coupled semiconductor devices'*, Bell systems tech. 587, 1970.

[67] Kosonocky W. F., Carnes J. E. *'Two phase charge-coupled shift registers'*, Digest of Tech. Papers, IEEE International Solid State Circuits Conference. 132, Feb 1972.

[68] Jorden P. Private Communication 1989.

[69] Intel. *'8255A/8255A-5 Programmable Peripheral Interface'*. 8-85 - 8-105.

[70] Hawkins, G. A., Rivaud, L., Kyan, J. *'Characterization of dark current in CCD Imagers'*, IEDM, 84, 556-559, 1984.

[71] Allan J. Private Communication 1991.

[72] Aikens, R., Duncan, R. et. al. *'A New CCD with ultra low dark current and high dynamic range'*, Pre-Publication Document, 164-169.

[73] Chen, J. Y., Viswanathan, C.R. *'Geometry Dependence of Dark Current in CCD'*, IEEE Trans. Electron. Devices, 12, 1914-1917, Dec. 1984.

[74] Nicholls R. L., Teter W. D. *'Computer graphics for convex polyhedra: Hidden line removal and shading'*, Eng. Design Graphics Journal 52(3), 28-37, Fall 1988.

[75] ElGindy H., Avis D., Toussaint G. *'Applications of a two-dimensional hidden-line algorithm to other geometric problems'*, 31(3), 191-202, 1983.

[76] *'Z8410 Z80 DMA Direct Memory Access Controller'*, Zilog Prod. Spec., 27-43, Sept. 1983.

[77] Ostberg O., Shahnavaz H., Stenberg R. *'CRT Flicker and Scan-Line Direction'*, 8(2), 75-78, April 1987.

[78] Guekos G., Ulmi R. *'Room illumination and CRT/Flicker In visual Display Terminals'*, 2(1), 6-11, July 1983.

[79] Hilbert, D. *'Ueber stetige Abbildung einer linie auf ein Flachenstuck'*, Math. Ann. 38, 459-460, 1891.

[80] Peano, G. *'Sur une courbe qui remplit totute une aire plaine'*, Math. Ann. 36, 157-160, 1890.

[81] Stevens, R. et al. *'Data ordering and compression of multispectral images using the peano scan'*, IEE International Conference on Electronic image processing, No. 214, 1980.

[82] Cole, A. J. *'Compaction techniques for raster scan graphics using space filling curves'*, Computer J., 30(1), 87-92, 1987.

[83] Cole A. J. *'Murray polygons as a tool in raster scan graphics'*, Pre-publication Document, 1988.

[84] Maitland A., Hirst P. F. *'Security System for Visual Display Units and Personal Computers'*, Jan. 1989.

[85] Lorraine Electronics. *'The Company and The Products'*.

[86] Turn, R. *'Advances In Computer Security vols. I, II, III'*, ISBN 0-890060096-7, 0-89006-156-4, 0-90006-315-X, 1983, 1984, 1988.

[87] Graf, R.F., Sheets, W. *'Video Scrambling & Descrambling for satellite & cable TV'*, ISBN 0-672-22499-2, 1986.

[88] Cornwall, H. *'Data Theft. Computer fraud, industrial espionage & information crime'*, ISBN 0-435-90265-9, 1987.

[89] *'Low-Cost Device Shields VDUs From Data Eavesdroppers'*, Comput. Fraud & Secur. Bull, 8(9), 8-9, July 1986.

[90] Marshall R. C. *'VDU Security device to prevent Unauthorized, Remote Decoding'*, July 1989.

[91] Datel Intersil. *'Microprocessor Compatible Double-Buffered D/A Converters DAC-608, DAC-610, DAC-612'*, 1989.

[92] Datel Intersil. *'Wide Bandwidth, FET Input Monolithic Operational Amplifier AM-410 & AM-411 SERIES'*, 314C-317C.

[93] Datel Intersil. *'Low Cost, 10 Bit Monolithic Digital-to-Analog Converter DAC-IC10B Series'*, 314C-317C, Oct 1979.

[94] Adams W. O. *'Magnetic Deflection Yokes'*, Proc. of the SID., 24/4, 363-367, 1983.

[95] National Semiconductor Corporation. *'LM 12(L/C/CL) 150W Op Amp'*, 2-272 - 2-285.

[96] National Semiconductor Corporation. *'LH0101/LH0101C, LH0101A/LH0101AC Power Operational Amplifier'*, 2-214 - 2-225.

[97] Agnew, G.B. *'Random Sources for Cryptographic Systems'*, 77-81.

[98] *'A Simple Algorithm for Fast Real-Time Generation of Pseudorandom Poisson Integers with Rapidly Varying Means'*, Proc. IEEE, 2088, 1969.

[99] Nakamura, Shogo. *'A Method Of Generating a Random Signal Using Operational Amplifiers'*, Proc. IEEE, 651-653, May 1974.

[100] Castanie, F. *'Generation of Random Bits with Accurate and Reproducible Statistical Properties'*, Proc. IEEE, 807-809, 1978.

[101] Box, G.E.P., Muller, M.E. Ann. Math. Statist. 29, 610-611, 1958.

[102] Golder, E.R., Settle, J.G. Appl. Statist., 25, 12-20, 1976.

[103] Encyclopedia of Statistical Sciences, Wiley, 9, 355-359, 1988.

[104] Encyclopedia of Statistical Sciences, Wiley, 9, 348-349, 1988.

[105] Encyclopedia of Statistical Sciences, Wiley, 1, 389-397, 1982.

[106] Hastings C. Jr. *'Approximations for digital computation'*, Princeton University Press. Princeton N.J. 1955.

[107] Abramowitz M., Stegun A. I. *'Handbook of Mathematical Functions'*, 932.

[108] Hawkes, A.G. *'Approximating the Normal tail'*, The Statist. 31(3), 231-236, 1982.

[109] Lesurf J. C. G. *'Chaos in electronics'*, Elecr. World & Wireless World. 97(1664), 467, 1991.

[110] Intel. *'27C020 2m (256k × 8) CMOS EPROM'*, May 1990.

[111] Micro Power Systems. *'CMOS 11-Bit Monolithic A/D Flash Converter MP7685'*. 2-68 - 2-74.

[112] Jorden P. R., Thorne D. J., Van Breda I. G. *'Royal Greenwich Observatory (RGO) charge-coupled device (CCD) camera'*, S.P.I.E. 331, 87-95, 1982.

[113] McGuire T. *'The Kitt Peak CCD Camera System'*, Pub. of the Astron. Soc. of the Pacific 95, 919-924, 1983.

[114] Hlivak R. J., Pilcher C. B., et at. *'The Galileo Institute for Astronomy (IFA) charge-coupled device (CCD) system'*, Proc. S.P.I.E. . 331, 1982.

[115] Mackay C. D. *'Drift Scan Observations with a Charge Coupled device (CCD)'*, S.P.I.E. 331, 146-150, 1982.

[116] Walker G. A. H., Johnson R. et al.*'The CFHT CCD Detector'*, Pre publication document.

[117] *'CCD05-30 Series Scientific Image Sensor'*, EEV Limited, (1), June 1990.

[118] *'CCD 111 256-Element Line Scan Image Sensor'*, Fairchild Weston Schlumberger, 1986.

[119] Livingston, W.C. *'Diode Arrays - A Review'*, Kitt Peak Nat. Obs., 22-1 - 22-13.

[120] Latham D.W. *'Spectroscopy with Photon-counting reticons and solid state imagers for astronomy'*, Proc. IAU colloquium 67, Sept 1982.

[121] Meyer, H.J., Schmidt, K.H., Rosenbauer, H. *'Reticon Detector Electronics for the Halley Multicolor Camera on the Giotto Space Mission'*, Academic Press, ISBN 0-12-014664-9, 223-230, 1985.

[122] Mackay, C.D. *'Low Noise CCD System Development at Cambridge'*, Conference paper.

[123] Geary, J.C. *'A floating gate preamplifier design for reticon diode arrays'*, S.P.I.E., 172, 82-84, 1979.

[124] Wood, B.C. *'Preamplifier and clock drivers for the University of California at Los Angles reticon spectrometer'*, S.P.I.E., 331, 338-356, 1982.

[125] Wamsteker W. *'Catching all the Photons: the CCD'*, Messenger, 13, 10-11, 1978.

[126] *'Peltier Cooled Package'*, EEV Limited, A44-1A-Peltier, Aug 1990.

[127] Bailey, P., Pool, P. *'The P88000 Series of Large Area CCDs for Visible, Near Infra-red, and X-RAY Scientific imaging Applications'*, EEV Publication 1191-23, Sept. 1989.

[128] *'UV Sensitive Coating for CCD Imagers'*, EEV Limited, May 1990.

[129] Lehmann, M.S., Kuhs, W.F., et. al. *'On the Use of Small Two-Dimensional Position-Sensitive Detector in Neutron Diffraction'*, J. Appl. Cryst., 22, 562-568, 1989.

[130] Lemonier, M., Richard, J.C., et. al. *'Photon-in and Electron-in CCD Arrays for Image read-out tubes'*, IEEED Conf., 253, 74-77, 1985.

[131] Zucciuno, P., Long, D. et. al. *'Evaluation of RCA thinned buried channel charge-coupled devices (CCDs) for Scientific applications'*, S.P.I.E., 290, 174-176, 1981.

[132] *'Thinned CCD Sensors for Back Illumination P86230/T Series'*, EEV Limited, Feb 1990.

[133] *'EEV CCD Imaging III'*, EEV U.K., 1987

[134] Castanie, F. *'Charge-coupled devices and applications'*, RCA, April 1981

[135] Thorne, D.J, Jorden, P.R. et. al. *'Laboratory and astronomical comparisons of RCA, GEC and Thomson CCDs'*, S.P.I.E., 627, 530-542, 1986.

[136] Mortara, L., Fowler, A. *'Evaluation of charge-coupled device (CCD) performance for astronomical use'*, S.P.I.E., 290, 28-33, 1981.

[137] Kristian J., Blouke, M. *'Charge-coupled Devices in Astronomy'*, Scientific American, 48-56, Oct. 1982.

[138] Hall, J.A. *'Applied Optics and Optical Engineering Vol. III'*, Academic Press Inc., Chap 8, ISBN 0-12-408608-X, 1980.

[139] Fraser, D.A. *'The Physics of Semiconductor Devices Fourth Edition'*, Oxford Science Publications, ISBN 0-19-851866-8, 147-150, 1986.

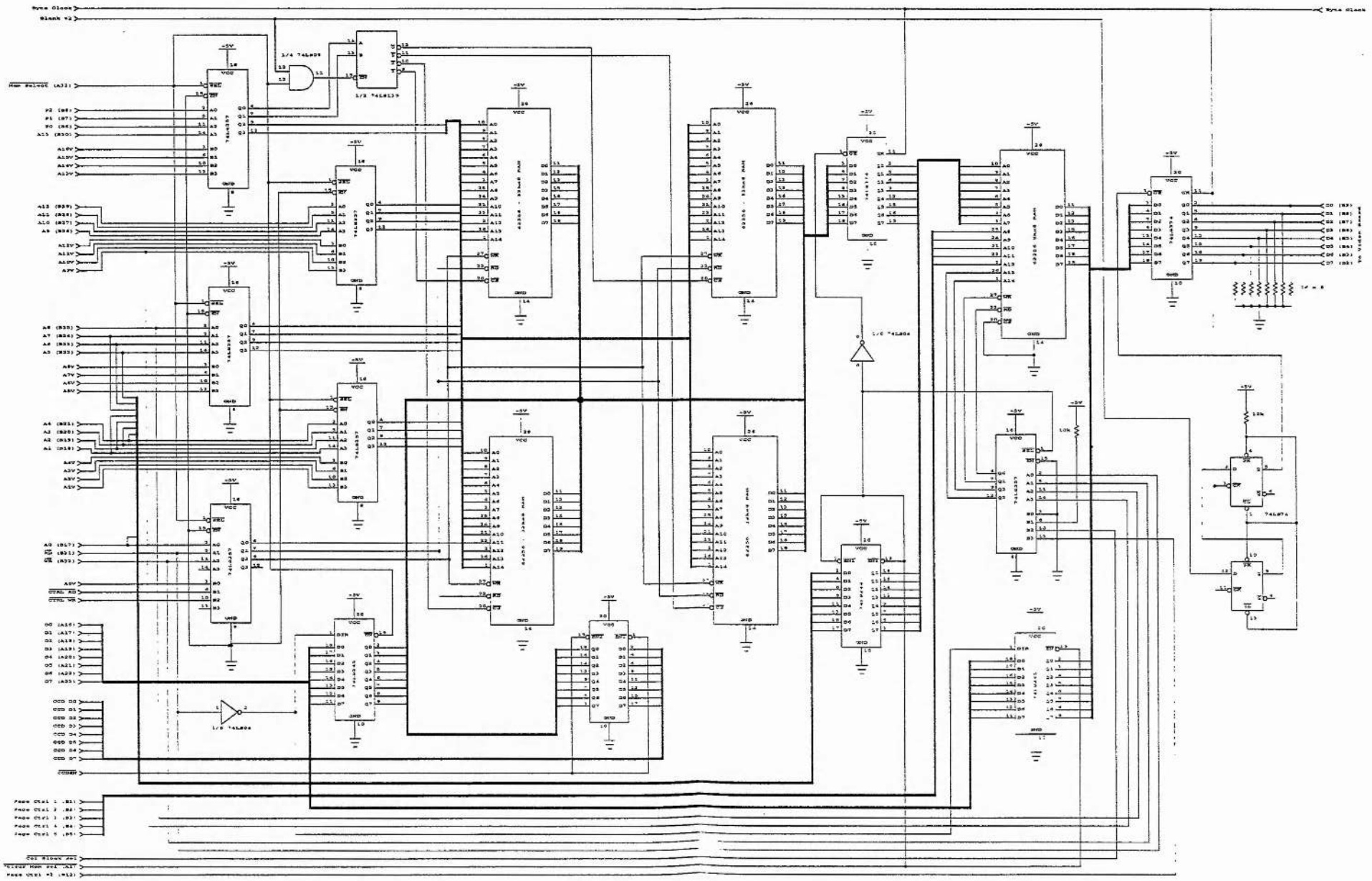


ILLUSTRATION I : Circuit diagram of the video memory board.

ILLUSTRATION I : Circuit diagram of the video memory board.

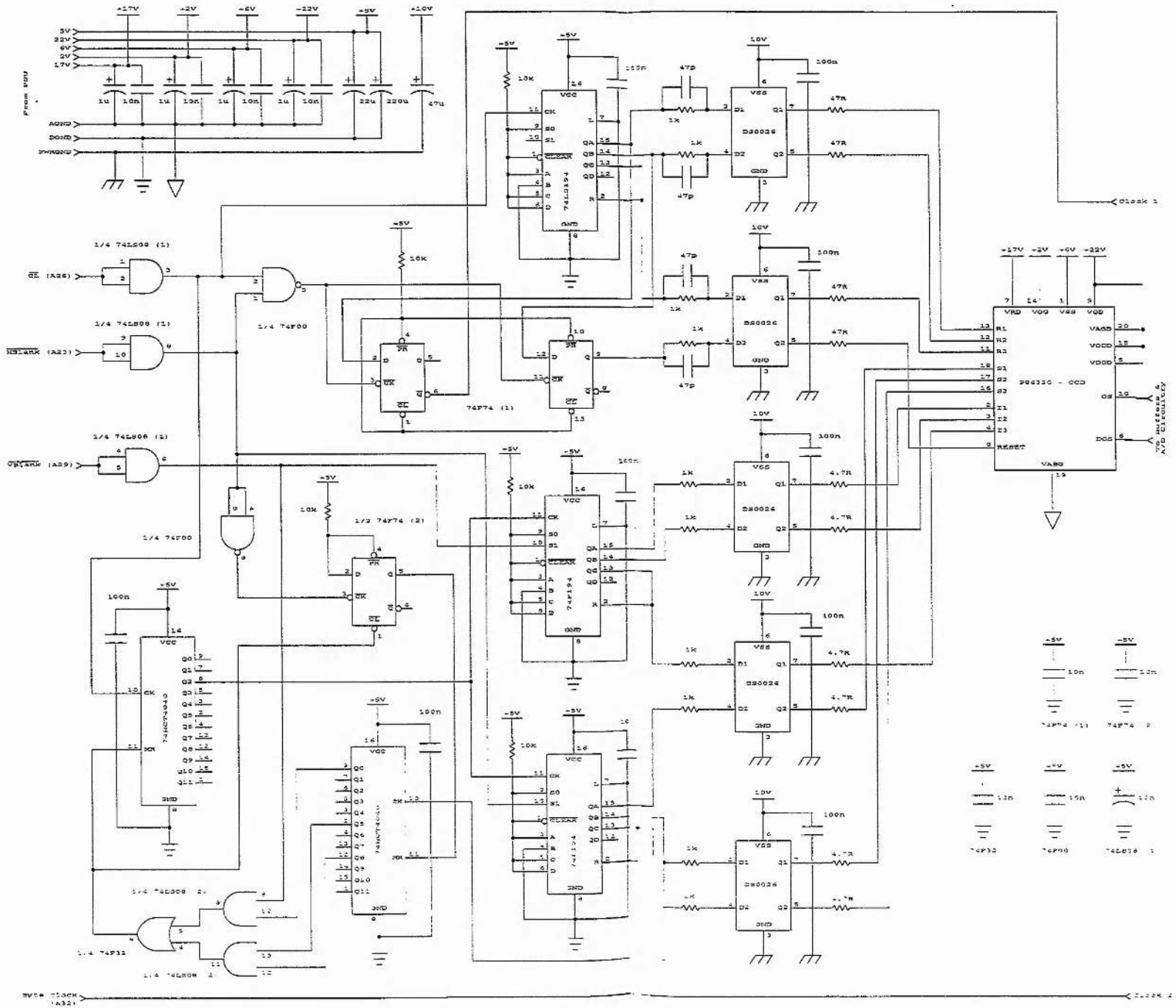


ILLUSTRATION II : Circuit diagram of the timing and video controller.

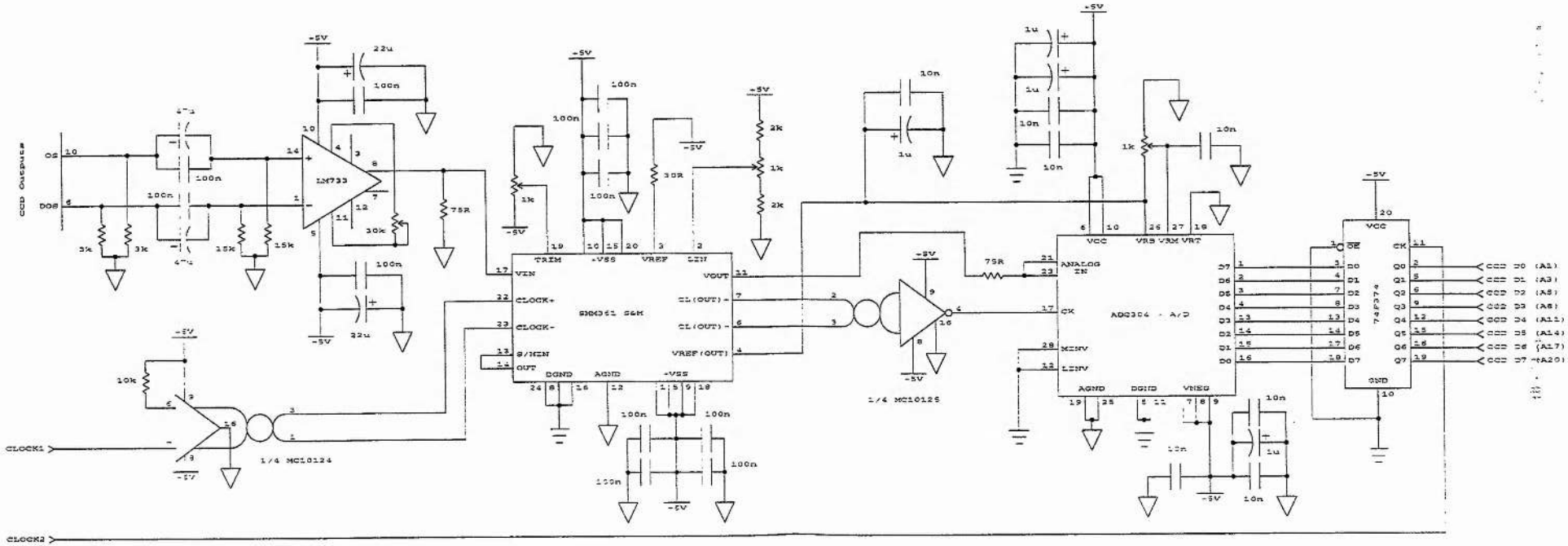


ILLUSTRATION III : Circuitry for ECL sampling and digitization.

ILLUSTRATION III : Circuitry for ECL sampling and digitization.

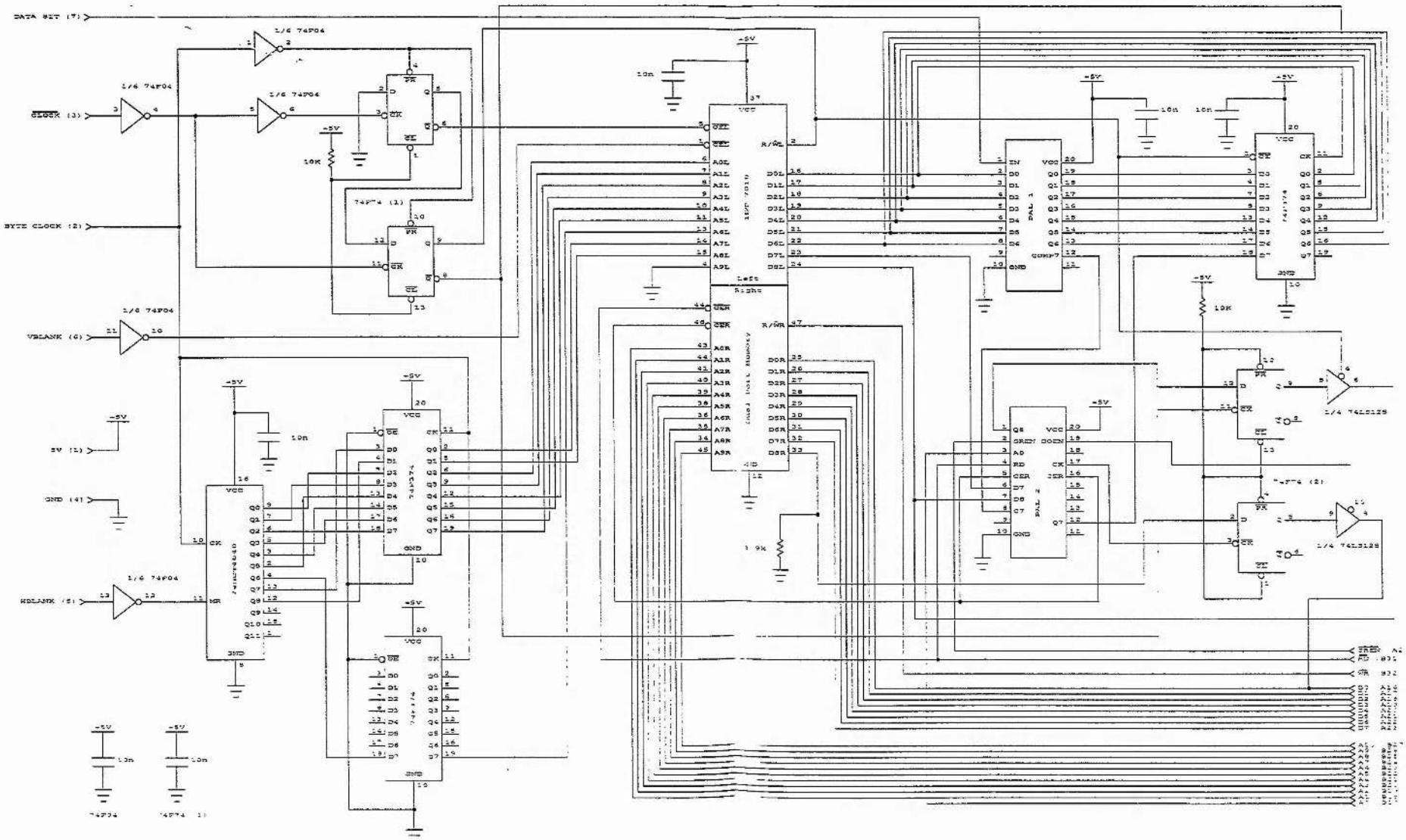


ILLUSTRATION IV : Circuit diagram of the vertical integration hardware.

ILLUSTRATION IV : Circuit diagram of the vertical integration hardware.

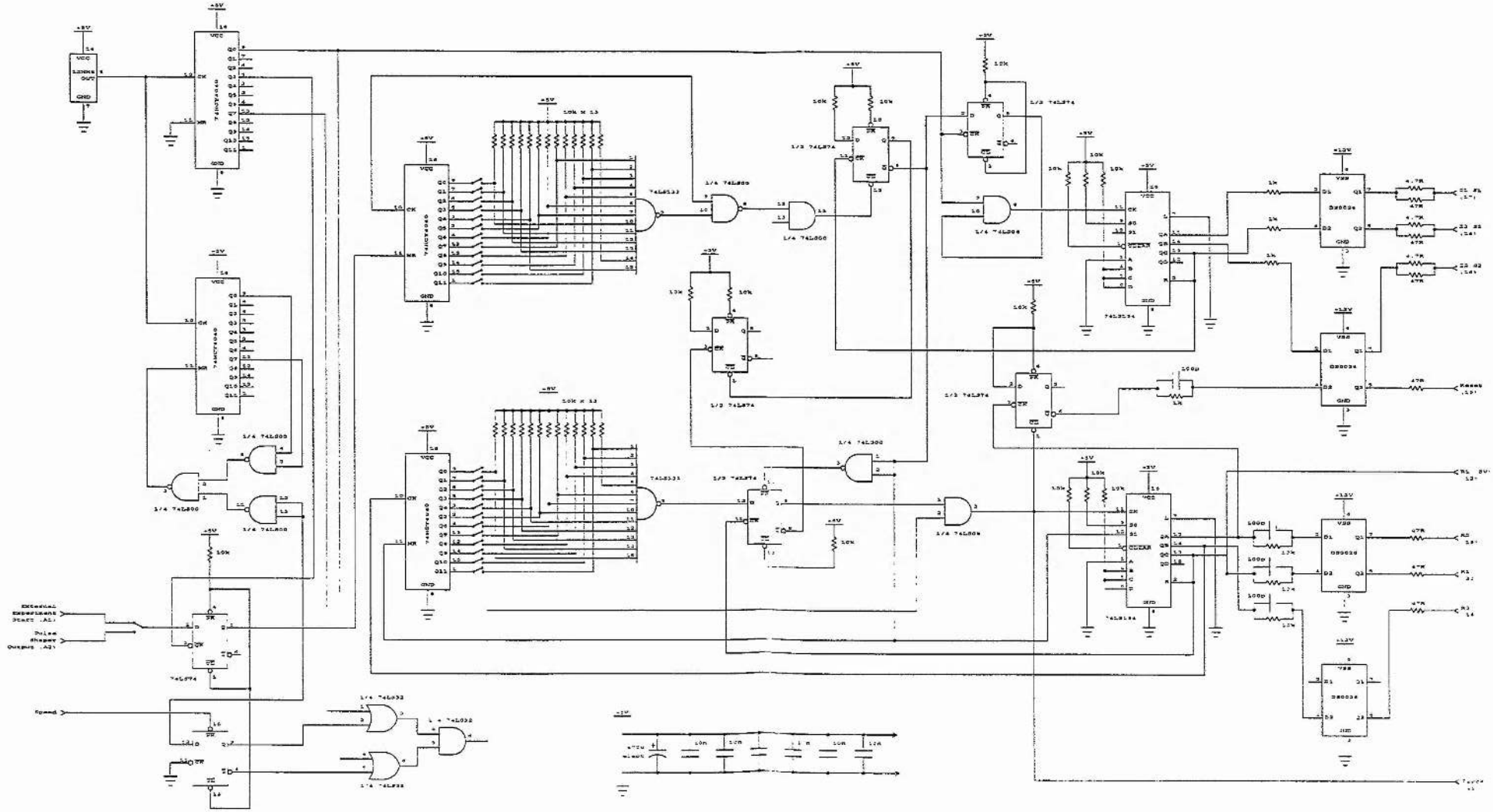


ILLUSTRATION V : Circuit diagram of the digital controller.

ILLUSTRATION V : Circuit diagram of the digital controller.

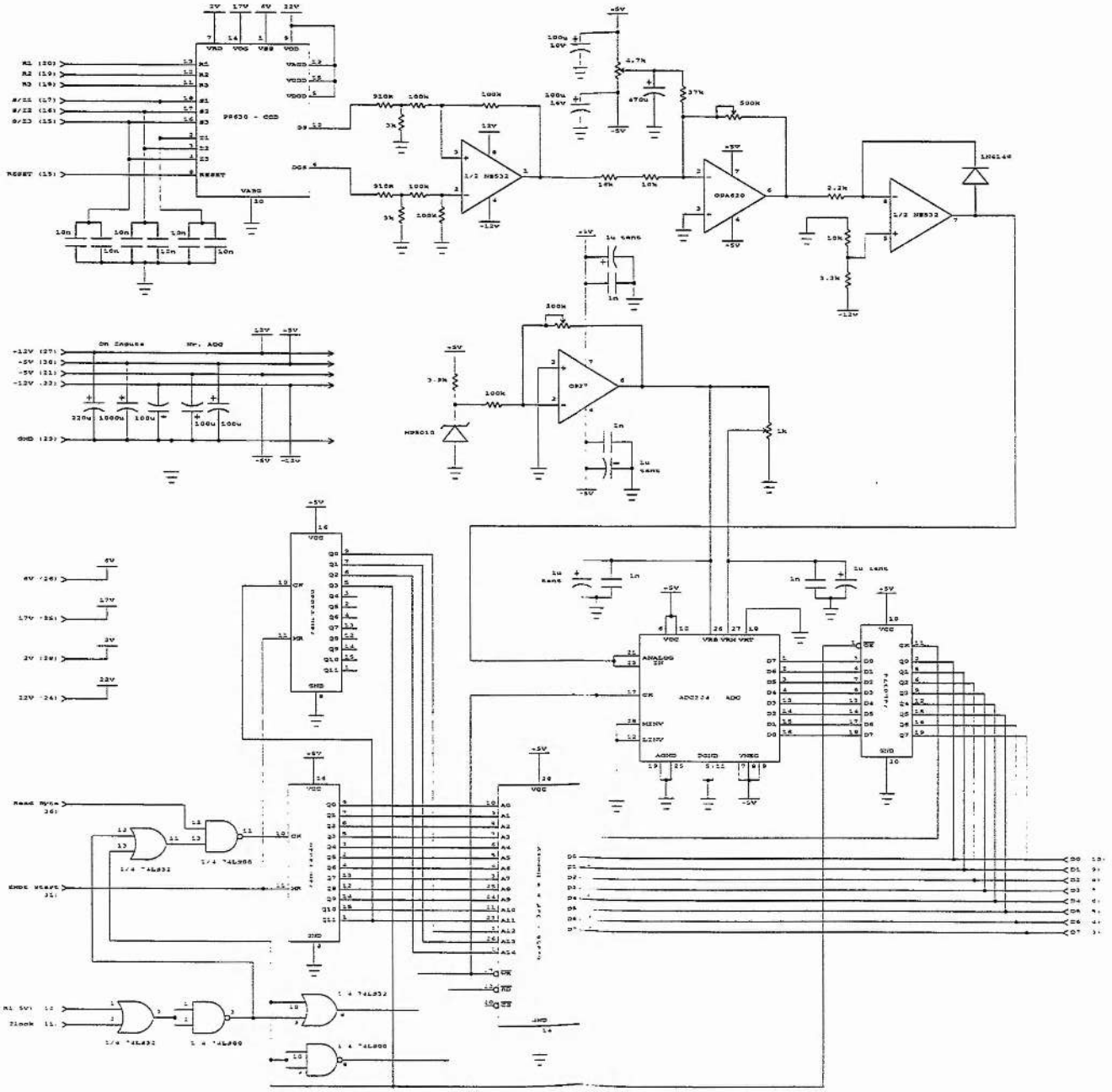


ILLUSTRATION VI : The analog driver and sampling stages.

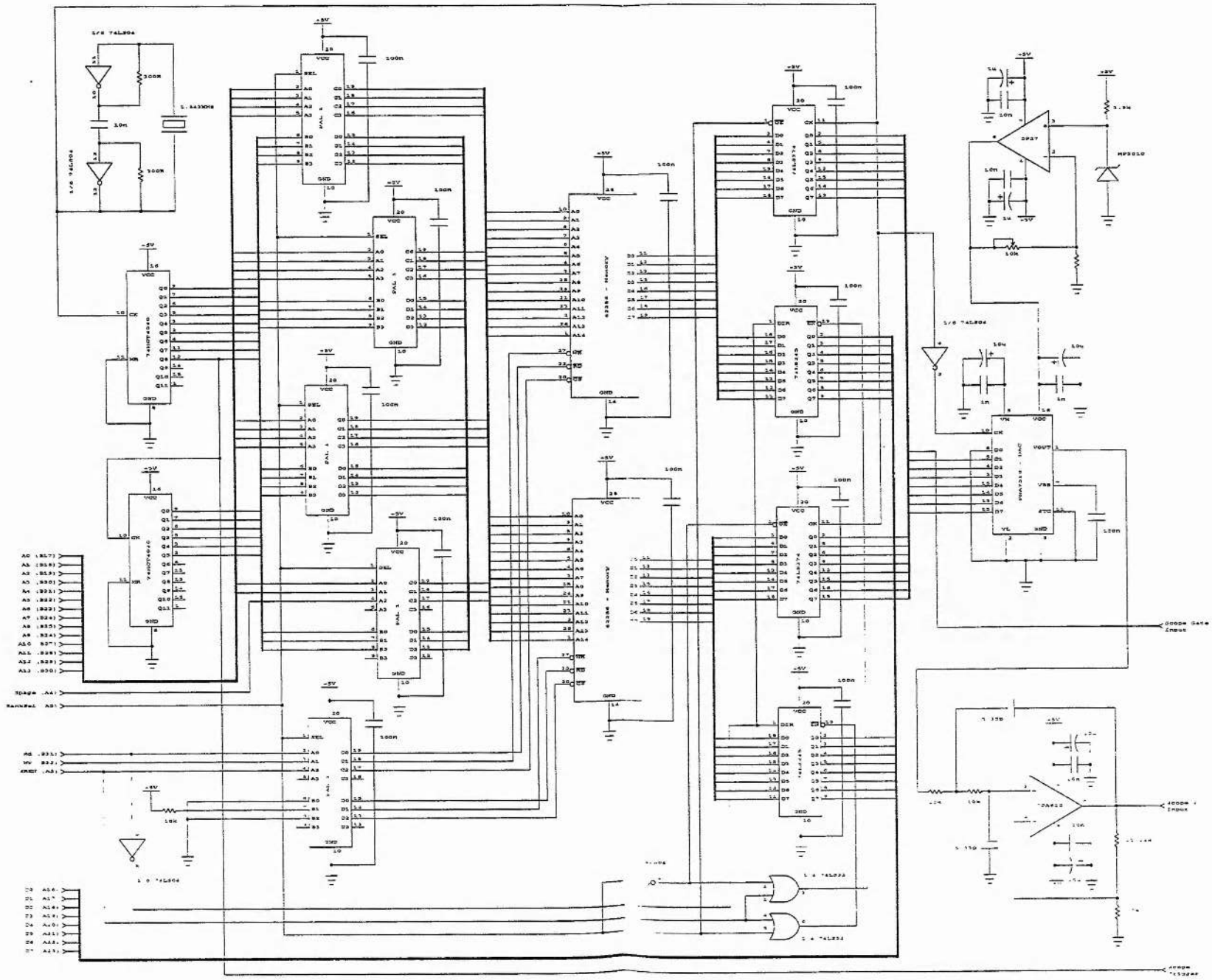


ILLUSTRATION VII : Circuit diagram of the basic surface generator.

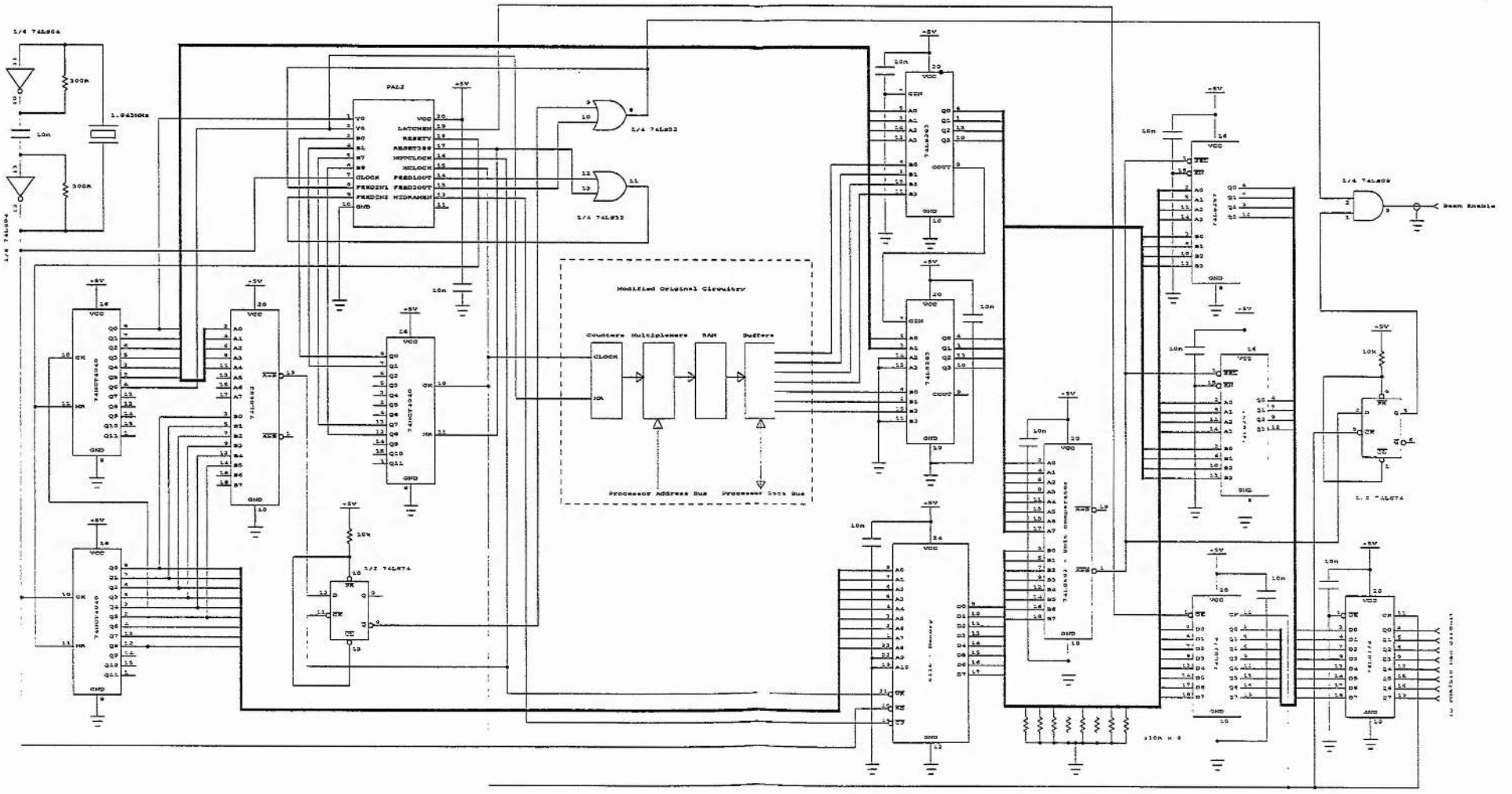


ILLUSTRATION VIII : Circuit diagram of the full surface generator.

ILLUSTRATION VIII : Circuit diagram of the full surface generator.